PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(F1) T-44'1 TD 4 4 CT 10 44 6	Τ	THE TITLE OF EXAMINITY TREATT (FCT)	
(51) International Patent Classification ⁶ :		(11) International Publication Number: WO 96/41883	
C12N 15/31, C07K 14/315, A61K 38/16		(43) International Publication Date: 27 December 1996 (27.12.96)	
(21) International Application Number: PCT/US	96/0964	(81) Designated States: CA, JP, MX, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL,	
(22) International Filing Date: 7 June 1996 (07.06.9	6) PT, SE).	
(30) Priority Data: 08/488,940 9 June 1995 (09.06.95)	τ.	Published With international search report	
9 Julie 1993 (05.00.93)	·	With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of	
(71) Applicant: PRESIDENT AND FELLOWS OF HA COLLEGE [US/US]; 17 Quincy Street, Cambrid 02139 (US).	amendments.		
(72) Inventor: REED, Guy, L.; 30 Vine Street, Winches 01890 (US).	ster, M	A	
(74) Agent: FRASER, Janis, K.; Fish & Richardson P Franklin Street, Boston, MA 02110-2804 (US).	P.C., 22	5	
·			

(54) Title: PLASMIN-RESISTANT STREPTOKINASE

(57) Abstract

The invention features modified streptokinase (SK) molecules which are resistant to plasmin cleavage including a recombinant fusion protein in which the amino terminus of SK was blocked with a peptide, a recombinant fusion protein in which an amino-terminal deleted SK was blocked with a peptide, and a mutated SK in which plasmin-cleavage sites were altered to render those sites resistant to enzymatic cleavage.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
ΑT	Austria	GE	Georgia *	MX	Mexico
ΑU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgystan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic	SD	Sudan
CF	Central African Republic	*****	of Korea	SE	
CG	Congo	KR	Republic of Korea	SG	Sweden
CH	Switzerland	KZ	Kazakhstan	IS	Singapore
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovenia
CM	Cameroon	LK	Sri Lanka	SN.	Slovakia
CN	China	LR	Liberia	SZ	Senegal
CS	Czechoslovakia	LT	Lithuania		Swaziland
CZ	Czech Republic	LU	Luxembourg	TD	Chad
DE	Germany	LV	Latvia	TG	Togo
DK	Denmark	MC	Monaco	TJ TT	Tajikistan
EE	Estonia	MD	Republic of Moldova	TT	Trinidad and Tobago
ES	Spain	MG	-	UA	Ukraine
FI	Finland	ML	Madagascar Mali	UG	Uganda
FR	France			US	United States of America
GA	Gabon	MN	Mongolia	UZ	Uzbekistan
GA	Gubon	MR	Mauritania	VN	Viet Nam

PLASMIN-RESISTANT STREPTOKINASE Background of the Invention

Streptokinase (SK), isolated from Group C 5 streptococcus, is used as a plasminogen activator to accelerate the lysis of the coronary thrombi that cause heart attacks. However, SK is by itself inert and must combine with human plasminogen to form a catalyticallyactive SK-plasminogen activator complex (SK-PAC) which 10 cleaves substrate plasminogen molecules. Studies of proteolytic fragments of SK and recombinant truncation mutants have defined regions of SK which are important for binding interactions with plasminogen in the construction of the activator complex. Through undefined 15 molecular interactions, an active site appears in the plasminogen moiety of the SK-PAC (Buck et al., 1968, J. Biol. Chem. 246:209-246). The SK-PAC then generates the active enzyme plasmin by clipping substrate plasminogen molecules at the Arg560-Val bond (Robbins et al., 1987, 20 In Colman et al., Hemostasis and thrombosis: basic principles and clinical practice, 2nd ed., Lippincott, Philadelphia, pp. 341-357).

Almost immediately after forming an active SK-PAC, the SK moiety is clipped to smaller molecular weight

25 forms (Siefring and Castellino, 1976, J. Biol. Chem.
251:3913-3920; Markus et al., 1976, J. Biol. Chem.
251:6495-6504). Cleavage of SK markedly reduces the catalytic activity of the activator complex (Markus et al., 1976, supra). Enzymatic studies of SK fragments

30 isolated after reacting with plasminogen at lower temperatures suggests that SK activity declines with progressive cleavage (Markus et al., 1976, supra).

Inactivation of SK in plasma as a result of plasmin cleavage reduces the therapeutic effectiveness of this plasminogen activation.

WO 96/41883 PCT/US96/09640

- 2 -

Summary of the Invention

The SK-derived compounds of the invention resist cleavage inactivation by plasmin, while retaining all or a substantial portion of the plasminogen-binding and catalytic activity of native SK. SK modified according to the invention is a more potent thrombolytic agent than native SK, and therefore, is a more useful therapeutic tool.

The invention features a compound containing (a) a plasminogen-binding fragment of SK and (b) a blocking group at the amino-terminus of the fragment. By the term "streptokinase" is meant an indirect plasminogen activator derived from streptococci. By the term "fragment" is meant a polypeptide containing less than or all of the native, full-length amino acid sequence of SK. SK may be recombinant or purified from streptococci, and the streptococci from which it is derived is preferably β-hemolytic. Alternatively, the streptokinase may be derived from an α-hemolytic streptococci. The streptococci from which SK is derived is preferably from Group C, e.g., Streptococcus equisimilus, however SK may also be derived from streptococci of Group A or Group G.

The compound is catalytically active and the rate of in vitro degradation in the presence of human

25 plasminogen is at least two times slower than the rate of native, full-length mature SK protein derived from Streptococcus equisimilus (nSK), i.e., the time required from the addition of SK to plasminogen to the disappearance of the band on a Western blot corresponding to the uncleaved nSK. For example, the time required for the disappearance of uncleaved nSK is about 2 min., whereas the time for the disappearance of modified SK ranges from 7 min. to greater than 20 min. By the term "catalytically active" is meant it possesses the ability

WO 96/41883 PCT/US96/09640

of SK to interact with plasminogen to form a SK-PAC capable of activating plasminogen to plasmin. By the term "degradation" is meant the process by which SK is reduced by plasmin cleavage into lower molecular weight fragments. The rate of degradation is measured by the disappearance of a full-length recombinant SK as detected by immunoblotting using anti-SK antibodies.

The compound preferably contains the amino acid sequence of SEQ ID NO:4. The blocking group of the

10 compound may be a peptide or a non-peptide blocking group which is located at the amino-terminus of the SK fragment. For example, a blocking group may be introduced by glycosylation or myristolization.

Preferably, the blocking group is least one heterologous amino acid; more preferably, the blocking group is a heterologous peptide of two or more amino acids; and most preferably, the blocking group is a fragment of or all of maltose binding protein (MBP). By the term

"heterologous" is meant an addition or substitution of one or more amino acids that is different from that found at the corresponding site in nSK.

The invention also includes a DNA, e.g., a DNA vector, containing a coding sequence which encodes the polypeptide portion of the compound of the invention, and a method of dissolving blood clots in a mammal by administering an effective amount of the compound. An effective amount of the compound is an amount which is effective in dissolving at least one blood clot in a patient.

The invention also features a plasminogen-binding fragment of SK which is catalytically active and the rate of in vitro degradation of which is at least two times slower than the rate of nSK in the presence of human plasminogen. The fragment preferably comprises at least 95% of the amino acid sequence of nSK; more preferably,

the fragment lacks one to five amino-terminal amino acids of nSK; more preferably, the fragment lacks one to ten amino-terminal amino acids; more preferably, the fragment lacks 1-24 amino acids. In a preferred embodiment, the fragment consists of amino acids 14-414 of nSK (SEQ ID NO:4). A fragment consisting of amino acids 14-414 of nSK (SEQ ID NO:4) may also contain at least one or more mutations selected from the group consisting of K36A, R45A, K51A, K59A, K61A, K147A, K333, R232A, K257A, K298A, K309A, R234A, R363A, K386A, K372A, R388A, R394A, and R401A.

The invention also includes an SK polypeptide which is catalytically active and the rate of in vitro degradation of which is at least two times slower 15 compared to the rate of nSK. By "polypeptide" is meant a chain of amino acids, regardless of length or posttranslational modification (e.g., glycosylation or phosphorylation). Preferably, the polypeptide consists of the amino acid sequence of nSK in which at least one 20 potential plasmin cleavage site has been mutated to render it resistant to plasmin cleavage. More preferably, the polypeptide contains one or more mutations selected from the group consisting of R10A, K36A, R45A, K51A, K59A, K61A, K147A, K333, R232A, K257A, 25 K298A, K309A, R234A, R363A, K386A, K372A, R388A, R394A, and R401A. Most preferably, the fragment is rSK5mut (SEQ ID NO:17), which contains the mutations, R10A, R36A, R45A, R51A, and R59A or rSK6mut, which contains the mutations R10A, R36A, R45A, R51A, R59A, and K386A (SEQ ID 30 NO:18). The invention also includes a DNA containing a coding sequence encoding the SK polypeptide of the invention and a method of dissolving blood clots in a mammal by administering to the mammal an effective amount of the SK polypeptide of the invention.

Other features and advantages of the invention will be apparent from the following description of the preferred embodiments thereof, and from the claims.

<u>Detailed Description</u>

5 The drawings will first be described.

<u>Drawings</u>

Fig. 1 is a photograph of a Western blot showing purification of a fusion protein with maltose binding protein linked to the amino terminus nSK (rSK), a fusion protein with MBP linked to the amino terminus of nSK in which the amino terminal 13 amino acids of nSK were deleted (rSKA14), and rSK5mut.

Fig. 2 is a graph showing plasminogen activation by nSK, rSK, and rSKA14.

Fig. 3 is a photograph of a Western blot showing plasmin cleavage of nSK.

Fig. 4 is a photograph of a Western blot showing plasmin cleavage of rSK (0-20 min.).

Fig. 5 is a photograph of a Western blot showing 20 plasmin cleavage of rSK_A14.

Fig. 6 is a photograph of a Western blot showing plasmin cleavage of rSK5mut.

Fig. 7 is a photograph of a Western blot showing comparative plasmin cleavage of rSK, rSKA14, nSK, and rSK5mut.

Modification of SK to render it resistant to degradation by plasmin

Within seconds, binding of SK to plasminogen to form SK-PAC, nSK is rapidly degraded at its amino
terminus by plasmin. Through the process of degradation, plasmin limits the thrombolytic efficacy of nSK.
According to the invention, SK can be modified in three different ways to render it resistant to plasmin

cleavage: (1) by blocking the amino terminus of nSK, e.g., with a heterologous peptide; (2) by deleting one or more amino terminal amino acids from nSK; and (3) by altering plasmin cleavage sites throughout nSK to render 5 them resistant to plasmin cleavage. In one example, a recombinant fusion protein was made in which the amino terminus of nSK was tethered in peptide linkage to MBP (rSK). In another example, a recombinant fusion protein was made in which the MBP was linked to the amino 10 terminus of nSK, the first 13 amino acids of which were In the third example, the nSK amino acid deleted. sequence was mutated at plasmin-cleavage sites to render those sites resistant to enzymatic cleavage, e.g., in the mutant rSK5mut, the K or R residue in five potential 15 plasmin cleavage sites were changed to A residues. In each case, plasmin cleavage yielded catalytically active plasmin cleavage products, but the rate of degradation was markedly reduced compared to that of nSK. addition to affecting the rate of degradation, mutation 20 of plasmin cleavage sites also significantly decreases the $\mathbf{K}_{\mathbf{m}}$ of amidolytic activity, which leads to greater catalytic efficiency.

Therapeutic Applications

The compounds of the invention can be used to lyse blood clots in a mammal. The compounds can be administered by any standard route including intraperitoneally, intramuscularly, subcutaneously, or intravenously. It is expected that the preferred route of administration will be intravenous. The compounds can be administered systemically to the bloodstream as well as locally within the blood vessel at the site of clot formation. Since the compounds of the invention are timed-release, they can be administered in a single dose rather than by continuous infusion.

As is well known in the medical arts, dosages for any one patient depends on many factors, including the patients general health, sex, size, body surface area, age, as well as the particular compound to be 5 administered, time and route of administration, and other drugs being administered concurrently. Dosages for the compounds of the invention will vary, but a preferred dosage for administration to human patients is approximately 20,000 units per kg of body weight (units 10 of SK are defined in Bulletin. World. Health. Org., 1965, 33:235). Determination of correct dosage for a given application is well within the abilities of one of ordinary skill in the art of pharmacology. Optimal dosage may be adjusted according to the condition of the 15 patient and response of the patient to therapy.

EXAMPLE 1: Modification of the amino terminus of streptokinase modulates the appearance of the active site in the SK-PAC

To examine the functional role of the amino

20 terminus of SK in the SK-PAC, the amino terminus of SK

was recombinantly modified by partial deletion of aminoterminal amino acids or by tethering of the amino
terminus with a blocking group, e.g., a heterologous
peptide. Functional activity of the modified SK was

25 evaluated by measuring (1) the rate of plasminogen
activation by SK-PAC, (2) the amidolytic activity of the
SK-PAC, and (3) the plasmin-mediated degradation of SK in
the SK-PAC.

Cloning, Expression and Purification of Streptokinase.

30

The SK gene (Malke et al., 1985, Gene 34:357-362) was cloned from *Streptococcus equisimilis* by the polymerase chain reaction (PCR), sequenced (U.S. Biochemicals, Cleveland, Ohio; Sanger et al., 1977, Proc.

Natl. Acad. Sci USA 74:5463) and subcloned into the pMAL vector for bacterial expression (New England Biolabs, Beverly, MA) using known methods, e.g., Reed et al., 1993, J. Immunol. 150:4407-4415; Reed et al., 1993, 5 Circulation 88:Abstract I-615). The expressed SK gene formed a fusion protein with maltose binding protein at its amino terminus (rSK). Restriction digestion of the SK gene with Hinc II removed the nucleotides encoding the amino terminal 13 amino acids of SK to produce deletion 10 mutant, rSKA14. These recombinant SK fusion proteins were purified by affinity chromatography on an amylose resin (New England Biolabs, Beverly, MA) as described by the supplier. The purity of the recombinant SK fusion proteins was assessed by SDS-PAGE (Laemmli, 1970, Nature 15 227:680-685). For some experiments, the SK fusion proteins were cut with factor Xa (Maina et al., 1988, Gene 74:365) and the MBP portion of the fusion protein removed by affinity chromatography on an amylose resin.

After purification, the relative concentrations of 20 the recombinant SKs were determined by comparative radioimmunoassay (RIA) using anti-SK monoclonal antibodies. Wells of a microtiter plate were coated with various concentrations of nSK (0, 2.5, 5, 10, 20, and 40 μ g/mL) or different dilutions of the recombinant SKs, 25 rSKA14 and rSK5mut. After nonspecific binding sites had been blocked with 1% bovine serum albumin, anti-SK monoclonal antibodies were added to each well in duplicate. After a 1-h incubation, the wells were washed and probed with ^{125}I goat anti-mouse antibody (Cappel 30 Organon Teknika, Durham, NC) for 1 h. After another wash, the amount of bound antibody was determined by gamma counting. A standard curve relating antibody binding (cpm) to nSK concentration was derived and the concentration of each recombinant SK was determined by

35 reference to the standard curve.

Plasminogen Activation by recombinant SKs.

Studies of the time-related activity of different SKs were carried out by mixing Glu-plasminogen (333 nM; American Diagnostica, Greenwich, CT) in a quartz cuvette with S2251 (0.5 mM; H-D-valyl-L-leucyl-L-lysine-p-nitroanilide dihydrochloride, Chromogenix, Sweden) at 21°C or 37°C and then adding purified nSK, rSK, or rSKA14 (16.7 nM). Absorption at 405 nanometers was continuously monitored in a Hewlett-Packard diode array spectrophotometer.

Active Site Titration

The development of an active site in the SK-PAC was monitored using standard methods. Plasminogen (8.5 μg; Sigma, St. Louis, MO) was added to a quartz cuvette containing 2 ml of filtered buffer (50 mM, 100 mM NaCl, pH 7.4) and 1 mM of the fluorogenic substrate 4-methylumbelliferyl p-guanidinobenzoate (Sigma, St. Louis, MO) thermostatically maintained at 25°C. The emission at 445 nanometers (excitation at 365 nanometers) was continuously monitored in a Hitachi 2000 fluorescence spectrophotometer. After ~200 seconds of observation, rSK was added, and the reaction was recorded for a total of 2000 seconds.

Kinetic Assays of the SK-PAC

The amidase kinetic parameters of nSK, rSK and rSKA14 were studied using a paranitroanilide substrate (S2251, H-D-valyl-L-leucyl-L-lysine-p-nitroanilide dihydrochloride, Chromogenix, Sweden) using known methods, e.g., Wohl R. et al., 1980., Biochim. et

Biophys. Acta 745:20-31). The recombinant SK proteins and Glu-plasminogen were mixed together and incubated for 5 min. (nSK and rSK) or 20 min (rSKA14) at 37°C. The mixture was then transferred to a quartz cuvette

WO 96/41883 PCT/US96/09640

containing assay buffer (50 mM Tris, 100 mM NaCl, pH 7.4) and various concentrations of S2251 (100-800 μ M) added. The cuvette was thermostatically regulated at 37°C. The change in absorbance was monitored at 404 nM for 10 min. 5 at 37°C, and the data were transformed to Linewaever-Burke plots to determine the K_m and V_{max} .

Studies of the degradation of SK by plasmin

The time-related proteolysis of nSK, rSK, rSKa14, and rSK5mut was studied by immunoblotting. nSK (1 μ g) or 10 recombinant SKs (2 μ g) were mixed together with purified human Glu-plasminogen (40 μ gs; American Diagnositica, 98% Glu-type plasminogen) for 0-20 min. The amount of human plasminogen present is typically in excess of the amount of SK. At various time points, an aliquot (5 μ l) was 15 removed and plunged into boiling water to stop the The samples were then electrophoresed on 10% SDS-polyacrylamide under reducing conditions and electrophoretically transferred to polyvinylidene difluoride membranes (Millipore, Bedford, MA). 20 Nonspecific binding sites were blocked with 5% nonfat The blots were incubated with pooled milk for 1 hr. monoclonal antibodies specific for SK overnight at 4°C. The blots were washed and incubated for 1 hr. with $^{125}\mathrm{I-}$

goat antimouse antibody (~1,000,000 cpm; Cappel Organon Teknika, Durham, NC) which had been labelled using the Iodogen labelling method known in the art. After washing, the blots were exposed to Kodak X-O-mat film (Rochester, NY) at -70°C.

Amino-terminal modification of SK

SK was produced as a fusion protein with MBP at its amino terminus (rSK), the amino acid sequence of which is shown in Table 1. A mutant lacking the first 13 amino acids of SK was also produced as a fusion protein (rSKA14), the amino acid sequence of which is shown in Table 2. The amino acid sequence of nSK is shown in Table 3, and the amino acid sequence of SKA14 is shown in Table 4. The sequence of both rSK and rSKA14 suggested that they could be cleaved at the fusion protein junction by factor Xa. The production of the rSK proteins in E. coli was induced by IPTG. Recombinant SK proteins were purified from bacterial lysates by affinity chromatography. As shown in Fig. 1, the proteins migrated at the predicted molecular size (rSK: 89 kDa, rSKA14: 87 kDa).

Table 1: rsk

MKTEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDG PDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEAL 20 SLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFK YENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTIN **GPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYL** LTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWY AVRTAVINAASGRQTVDEALKDAQTNSSSVPGRGSIEGRIAGPEWLLDRPSVNNSQL 25 VVSVAGTVEGTNQDISLKFFEIDLTSRPAHGGKTEQGLSPKSKPFATDSGAMSHKLE KADLLKAIQEQLIANVHSNDDYFEVIDFASDATITDRNGKVYFADKDGSVTLPTQPV **QEFLLSGHVRVRYKEKPIQNQAKSVDVEYTVQFTPLNPDDDFRPGLKDTKLLKTLAI** GDTITSQELLAQAQSILNKNHPGYTIYERDSSIVTHDNDIFRTILPMDQEFTYRVKN REQAYRINKKSGLNEEINNTDLISEKYYVLKKGEKPYDPFDRSHLKLFTIKYVDVDT 30 NELLKSEQLLTASERNLDFRDLYDPRDKAKLLYNNLDAFGIMDYTLTGKVEDNHDDT NRIITVYMGKRPEGENASYHLAYDKDRYTEEEREVYSYLRYTGTPIPDNPNDKNNSQ LVVSVAGTVEGTNQDISLKFFEIDLTSRPAHGGKTEQGLSPKSKPFATDSGAMSHKL EKADLLKAIQEQLIANVHSNDDYFEVIDFASDATITDRNGKVYFADKDGSVTLPTQP VQEFLLSGHVRVRYKEKPIQNQAKSVDVEYTVQFTPLNPDDDFRPGLKDTKLLKTLA

IGDTITSQELLAQAQSILNKNHPGYTIYERDSSIVTHDNDIFRTILPMDQEFTYRVK
NREQAYRINKKSGLNEEINNTDLISEKYYVLKKGEKPYDPFDRSHLKLFTIKYVDVD
TNELLKSEQLLTASERNLDFRDLYDPRDKAKLLYNNLDAFGIMDYTLTGKVEDNHDD
TNRIITVYMGKRPEGENASYHLAYDKDRYTEEEREVYSYLRYTGTPIPDNPNDK
5 (SEQ ID NO:1)

Table 2: rSK₁14

MKTEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDG PDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEAL SLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFK 10 YENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTIN GPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYL LTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWY AVRTAVINAASGRQTVDEALKDAQTNSSSVPGRGSIEGRNNSQLVVSVAGTVEGTNQ DISLKFFEIDLTSRPAHGGKTEQGLSPKSKPFATDSGAMSHKLEKADLLKAIQEQLI 15 ANVHSNDDYFEVIDFASDATITDRNGKVYFADKDGSVTLPTQPVQEFLLSGHVRVRY KEKPIQNQAKSVDVEYTVQFTPLNPDDDFRPGLKDTKLLKTLAIGDTITSQELLAQA QSILNKNHPGYTIYERDSSIVTHDNDIFRTILPMDQEFTYRVKNREQAYRINKKSGL NEEINNTDLISEKYYVLKKGEKPYDPFDRSHLKLFTIKYVDVDTNELLKSEQLLTAS ERNLDFRDLYDPRDKAKLLYNNLDAFGIMDYTLTGKVEDNHDDTNRIITVYMGKRPE 20 GENASYHLAYDKDRYTEEEREVYSYLRYTGTPIPDNPNDKNNSQLVVSVAGTVEGTN QDISLKFFEIDLTSRPAHGGKTEQGLSPKSKPFATDSGAMSHKLEKADLLKAIQEQL IANVHSNDDYFEVIDFASDATITDRNGKVYFADKDGSVTLPTQPVQEFLLSGHVRVR YKEKPIQNQAKSVDVEYTVQFTPLNPDDDFRPGLKDTKLLKTLAIGDTITSQELLAQ AQSILNKNHPGYTIYERDSSIVTHDNDIFRTILPMDQEFTYRVKNREQAYRINKKSG 25 LNEEINNTDLISEKYYVLKKGEKPYDPFDRSHLKLFTIKYVDVDTNELLKSEQLLTA SERNLDFRDLYDPRDKAKLLYNNLDAFGIMDYTLTGKVEDNHDDTNRIITVYMGKRP EGENASYHLAYDKDRYTEEEREVYSYLRYTGTPIPDNPNDK (SEQ ID NO:2)

Table 3: nSK

IAGPEWLLDRPSVNNSQLVVSVAGTVEGTNQDISLKFFEIDLTSRPAHGGKTEQGLS
PKSKPFATDSGAMSHKLEKADLLKAIQEQLIANVHSNDDYFEVIDFASDATITDRNG
KVYFADKDGSVTLPTQPVQEFLLSGHVRVRYKEKPIQNQAKSVDVEYTVQFTPLNPD
DDFRPGLKDTKLLKTLAIGDTITSQELLAQAQSILNKNHPGYTIYERDSSIVTHDND
IFRTILPMDQEFTYRVKNREQAYRINKKSGLNEEINNTDLISEKYYVLKKGEKPYDP

FDRSHLKLFTIKYVDVDTNELLKSEQLLTASERNLDFRDLYDPRDKAKLLYNNLDAF
GIMDYTLTGKVEDNHDDTNRIITVYMGKRPEGENASYHLAYDKDRYTEEEREVYSYL
RYTGTPIPDNPNDKNNSQLVVSVAGTVEGTNQDISLKFFEIDLTSRPAHGGKTEQGL
SPKSKPFATDSGAMSHKLEKADLLKAIQEQLIANVHSNDDYFEVIDFASDATITDRN
5 GKVYFADKDGSVTLPTQPVQEFLLSGHVRVRYKEKPIQNQAKSVDVEYTVQFTPLNP
DDDFRPGLKDTKLLKTLAIGDTITSQELLAQAQSILNKNHPGYTIYERDSSIVTHDN
DIFRTILPMDQEFTYRVKNREQAYRINKKSGLNEEINNTDLISEKYYVLKKGEKPYD
PFDRSHLKLFTIKYVDVDTNELLKSEQLLTASERNLDFRDLYDPRDKAKLLYNNLDA
FGIMDYTLTGKVEDNHDDTNRIITVYMGKRPEGENASYHLAYDKDRYTEEEREVYSY
10 LRYTGTPIPDNPNDK (SEQ ID NO:3)

Table 4: SK_{\(\Delta\)}14

NNSQLVVSVAGTVEGTNQDISLKFFEIDLTSRPAHGGKTEQGLSPKSKPFATDSGAM
SHKLEKADLLKAIQEQLIANVHSNDDYFEVIDFASDATITDRNGKVYFADKDGSVTL
PTQPVQEFLLSGHVRVRYKEKPIQNQAKSVDVEYTVQFTPLNPDDDFRPGLKDTKLL
15 KTLAIGDTITSQELLAQAQSILNKNHPGYTIYERDSSIVTHDNDIFRTILPMDQEFT
YRVKNREQAYRINKKSGLNEEINNTDLISEKYYVLKKGEKPYDPFDRSHLKLFTIKY
VDVDTNELLKSEQLLTASERNLDFRDLYDPRDKAKLLYNNLDAFGIMDYTLTGKVED
NHDDTNRIITVYMGKRPEGENASYHLAYDKDRYTEEEREVYSYLRYTGTPIPDNPND
KNNSQLVVSVAGTVEGTNQDISLKFFEIDLTSRPAHGGKTEQGLSPKSKPFATDSGA
20 MSHKLEKADLLKAIQEQLIANVHSNDDYFEVIDFASDATITDRNGKVYFADKDGSVT
LPTQPVQEFLLSGHVRVRYKEKPIQNQAKSVDVEYTVQFTPLNPDDDFRPGLKDTKL
LKTLAIGDTITSQELLAQAQSILNKNHPGYTIYERDSSIVTHDNDIFRTILPMDQEF
TYRVKNREQAYRINKKSGLNEEINNTDLISEKYYVLKKGEKPYDPFDRSHLKLFTIK
YVDVDTNELLKSEQLLTASERNLDFRDLYDPRDKAKLLYNNLDAFGIMDYTLTGKVE
25 DNHDDTNRIITVYMGKRPEGENASYHLAYDKDRYTEEEREVYSYLRYTGTPIPDNPN
DK (SEQ ID NO:4)

Functional activity of recombinant SKs

To compare the function of SK, rSK and rSKA14, the rate of plasminogen activation by these proteins was

sexamined at 21°C. nSK rapidly activated plasminogen with a minimal lag phase, i.e., less than 50 sec. (see Fig. 2). However, when expressed as a fusion protein, rSK showed a lag phase in plasminogen activation of

approximately 150 sec. (see Fig. 2). When expressed as a fusion protein lacking the amino terminal 13 amino acids, rSKA14 also showed a marked delay in time to activation os approximately 250 sec. (see Fig. 2). The lag phase refers to the time required for the reaction to the exponential phase of activity, e.g, full catalytic activity.

Plasmin cleavage products

Since nSK is known to be cleaved by plasmin after 10 formation of the SK-PAC, the rate of cleavage of rSK and rSKA14 was examined after various times of incubation with Glu-plasminogen. In these experiments, SK was mixed with an excess of plasminogen for various amounts of time and the resulting cleavage of SK was determined by 15 immunoblotting with monoclonal anti-SK antibodies. nSK was found to be rapidly degraded by plasmin within 30 secs to four lower molecular weight species, predominantly a ~36 kDa fragment (see Figs. 3 and 7). contrast, the degradation of rSK was slower, yielding a 20 fragment of 47 kDa (identical in size to nSK), first appearing at 1 min. A pattern of smaller SK fragments similar to that observed with nSK developed thereafter. After 5 min., a ~36 kDa SK fragment similar to that seen after nSK cleavage was found to be the major remnant from 25 rSK (see Figs. 4 and 7). Other lower molecular weight SK fragments, e.g., ~28 kDa, were also evident as cleavage products of nSK, and at later time points, of rSK. Plasmin cleavage products of rSKA14 are shown in Fig. 5.

Amino-terminal deletion of SK

The amino terminal 13 residues of SK are highly conserved among the SKs produced by different groups of streptococci. In addition, this region constitutes a major epitope for both murine and human antibodies

against SK. Removal of the amino-terminal 13 amino acids from nSK resulted in a further increase in the lag phase of plasminogen activation by rSKA14, as compared to rSK. This lag phase was marked at 21°C, but shortened

5 significantly when the temperature was raised to 37°C. Active site titration experiments indicate that removal of the amino terminus further delays the generation of the active site in the rSKA14.

Advantages of amino-terminally modified SK

At 37°C, and in vivo, nSK rapidly forms an active 10 site with plasminogen. The kinetics of this activation has been regarded as suboptimal for therapy because plasmin is rapidly activated in one large burst in vivo. To overcome the explosive activation of plasminogen, an 15 acylated SK-PAC (APSAC) made from SK and purified human plasminogen has been created in vitro (Ferres, 1987, Drugs 33 (Suppl. 3) 33). This approach permits APSAC to be given as a single bolus in vivo because continuous deacylation of the active site proceeds with a half-life 20 of 40 mins (Staniforth et al., 1983, Eur. J. Clin. Pharmacol. 24:751). A limitation of this approach is that the rate of appearance of the active SK-PAC is determined by the rate of deacylation and can not be otherwise modulated.

In contrast, recombinant modification of the amino terminus of SK, either by expression as a fusion protein, or by deletion of the amino terminus, can predictably alter the rate of active site generation. For example, the extent to which the rate of degradation is reduced compared to nSK is directly proportional to the number of deleted amino-terminal amino acids (up to 13 amino acids). Other advantages of the SK-derived compounds of the invention include a short half-life: 2-4 min.; safety: the compounds of the invention are not made from

human blood products; and cost-effectiveness: the compounds of the invention are recombinantly produced. The activity of the compounds is timed-released, therefore they can be administered in a single dose. The time required to achieve SK activity may also be modified depending on the number of amino-terminal amino acids removed from the nSK, i.e., length of time required is directly proportional to the number of amino acids deleted. In this manner, the timed-release activity of SK can be customized to suit the specific clinical application or patient to be treated. Thus, the compounds of the invention are improved clinical reagents because, using modified rSKs, an active SK-PAC can be generated at a rate consistent with best thrombolytic results.

EXAMPLE 2: Site-directed streptokinase mutants resist cleavage and degradation by plasmin

To examine the effects of cleavage on the activity of SK, site-directed mutations of R or L to A at putative plasmin or trypsin cleavage sites in the amino and carboxy terminus of SK were generated. The cleavage rate of these recombinant SKs were then examined. The catalytic function of rSKs with these specific mutations was also evaluated.

25 SK cloning and mutation by overlap extension

The SK gene was cloned from Group C Streptococcus equisimilis as described above. A series of mutations was performed in the amino terminus of SK to replace R or K residues with an A residue at putative plasmin cleavage sites. In addition, a single K to A mutation was constructed for K386 in the carboxy terminus of SK. PCR primers were used to produce site-directed mutations by the overlap extension method. For example, using nSK in

WO 96/41883 PCT/US96/09640

the pMAL vector as a template, PCR was performed using a primer corresponding to the mal E sequence of the pMALc vector and the SK 10 AS primer. At the same time the SK 10 S primer was used in a PCR reaction with a SK 36 AS 5 primer. The PCR products were purified on a low-melt agarose gel and used in an overlap PCR reaction. overlapped product was then further amplified using the mal E primer and the SK 36 AS primer. In a similar fashion, the primers were used to construct mutations at 10 the 45 and 51 position. The final overlap construct was between the 5' overlapped mutated SK segment containing the mutations at SK 10, 36, 45, and 51 and the segment from 51 to 127. This overlapped fragment was then ligated into the pMALc nSK, replacing the wild type 15 sequence, between restriction sites for KpnI and AflII. The SK 59 mutation was separately constructed and used to replace the wild type sequence between AflII and MunI. The mutation at residue 386 was similarly constructed and ligated into SK using a HindIII site. The mutated 20 pMALcSKs were sequenced to verify the desired mutations.

- 18 -

Table 5. Primers for Mutation by Overlap Extension

	<u>Primer Mutation</u> Restriction Site	Primer Sequence	
5	SK 10 S R->A	5'-GCTGCTAGACGCGCCATCTGTCAAC (SEQ ID NO:5)	HhaI
	SK 10 AS	5'-TGGCGCGTCTAGCAGCCACTCAG (SEQ ID NO:6)	
	SK 36 S K->A	5'-CAAGACATTAGTCTGGCCTTTTTTGAAATCG (SEQ ID NO:7)	HaeIII
10	SK 36 AS	5'-GGCCAGACTAATGTCTTGATTCG (SEQ ID NO:8)	
	SK 45 S R->A	5'-CGATCTAACATCGGCGCCTGCTCATGG (SEQ ID NO:9)	NarI
15	SK 45 AS	5'-CGCCGATGTTAGATCGATTTC (SEQ ID NO:10)	
	SK 51 S K->A	5'-GCTCATGGAGGCGCCACAGAGGGC (SEQ ID NO:11)	NarI
	SK 51 AS	5'-GGCGCCTCCATGAGCAGGTC (SEQ ID NO:12)	
20	SK 59 S K->A	5'-GCTTAAGTCCGGCCTCAAAACCATTTGC (SEQ ID NO:13)	HaeIII
	SK 59 AS	5'-TGAGGCCGGACTTAAGCCTTGCTC (SEQ ID NO:14)	
25	SK 386S K->A	5'-GCCGATCGATATACCGAAGAAGAACGAG (SEQ ID NO:15)	ClaI
		5'-TATCGATCGGCATCATAGGCTAAATGATAGC (SEQ ID NO:16)	

Plasmin-resistant SK site mutants

The following plasmin cleavage sites can be
mutated: R10A, K36A, R45A, K51A, K59A, K61A, K147A, K333,
R232A, K257A, K298A, K309A, R234A, R363A, K386A, K372A,
R388A, R394A, and R401A. Single mutants K59A, K386A,
were made, and the multiple mutant containing R10A, K36A,
R45A, K51A, and K59A (rSKmut5) was studied further.

Purification of rSK5mut is shown in Fig. 1. Multiple mutant rSK6mut is identical to rSK5mut with the addition of another mutation at a carboxy-terminal potential plasmin cleavage site. This mutant contains the

following mutations: R10A, K36A, R45A, K51A, K59A and k386A.

The plasmin-resistant SK site mutants produce catalytically-active plasmin cleavage products which are larger than those generated from nSK (see Figs. 6 and 7). The rate of degradation of rSK5mut is also slower than that of nSK (see Figs. 6 and 7).

Kinetic studies were performed to examine the catalytic activity of the site mutants. Table 6 shows the results from kinetic studies for rSK5mut and Gluplasminogen. These data show that mutation of plasmin cleavage sites significantly decreases the $K_{\rm m}$ of SK amidolytic activity leading to greater catalytic efficiency, and thus, greater therapeutic efficacy.

Table 6: Kinetic Parameters for recombinant SKs and Glu-Plasminogen

		(μM)	(S ^E 1)	$^{ ext{k}_{ ext{cat}}/ ext{K}_{ ext{m}}}_{(\mu ext{M}^{-1} ext{S}^{-1})}$
	nSK	248	56	0.226
20	rsk	152	42	0.276
	rSK∆14	533	51	0.096
	rSK5mut	77	52	0.675

Table 7: rSK5mut

MKTEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDG PDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEAL SLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFK 5 YENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTIN **GPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYL** LTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWY AVRTAVINAASGRQTVDEALKDAQTNSSSVPGRGSIEGRIAGPEWLLDAPSVNNSQL VVSVAGTVEGTNQDISLAFFEIDLTSAPAHGGATEQGLSPASKPFATDSGAMSHKLE 10 KADLLKAIQEQLIANVHSNDDYFEVIDFASDATITDRNGKVYFADKDGSVTLPTQPV QEFLLSGHVRVRYKEKPIQNQAKSVDVEYTVQFTPLNPDDDFRPGLKDTKLLKTLAI GDTITSQELLAQAQSILNKNHPGYTIYERDSSIVTHDNDIFRTILPMDQEFTYRVKN REQAYRINKKSGLNEEINNTDLISEKYYVLKKGEKPYDPFDRSHLKLFTIKYVDVDT NELLKSEQLLTASERNLDFRDLYDPRDKAKLLYNNLDAFGIMDYTLTGKVEDNHDDT 15 NRIITVYMGKRPEGENASYHLAYDKDRYTEEEREVYSYLRYTGTPIPDNPNDKNNSO LVVSVAGTVEGTNODISLKFFEIDLTSRPAHGGKTEOGLSPKSKPFATDSGAMSHKL **EKADLLKAIQEQLIANVHSNDDYFEVIDFASDATITDRNGKVYFADKDGSVTLPTQP** VQEFLLSGHVRVRYKEKPIQNQAKSVDVEYTVQFTPLNPDDDFRPGLKDTKLLKTLA IGDTITSQELLAQAQSILNKNHPGYTIYERDSSIVTHDNDIFRTILPMDQEFTYRVK 20 NREQAYRINKKSGLNEEINNTDLISEKYYVLKKGEKPYDPFDRSHLKLFTIKYVDVD TNELLKSEQLLTASERNLDFRDLYDPRDKAKLLYNNLDAFGIMDYTLTGKVEDNHDD TNRIITVYMGKRPEGENASYHLAYDKDRYTEEEREVYSYLRYTGTPIPDNPNDK (SEQ ID NO:17)

Table 8: rSK6mut

MKTEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDG PDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEAL SLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFK 5 YENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTIN GPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYL LTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWY AVRTAVINAASGRQTVDEALKDAQTNSSSVPGRGSIEGRIAGPEWLLDAPSVNNSQL VVSVAGTVEGTNQDISLAFFEIDLTSAPAHGGATEQGLSPASKPFATDSGAMSHKLE 10 KADLLKAIQEQLIANVHSNDDYFEVIDFASDATITDRNGKVYFADKDGSVTLPTQPV QEFLLSGHVRVRYKEKPIQNQAKSVDVEYTVQFTPLNPDDDFRPGLKDTKLLKTLAI GDTITSQELLAQAQSILNKNHPGYTIYERDSSIVTHDNDIFRTILPMDQEFTYRVKN REQAYRINKKSGLNEEINNTDLISEKYYVLKKGEKPYDPFDRSHLKLFTIKYVDVDT NELLKSEQLLTASERNLDFRDLYDPRDKAKLLYNNLDAFGIMDYTLTGKVEDNHDDT 15 NRIITVYMGKRPEGENASYHLAYDADRYTEEEREVYSYLRYTGTPIPDNPNDKNNSQ LVVSVAGTVEGTNQDISLKFFEIDLTSRPAHGGKTEQGLSPKSKPFATDSGAMSHKL **EKADLLKAIQEQLIANVHSNDDYFEVIDFASDATITDRNGKVYFADKDGSVTLPTQP** VQEFLLSGHVRVRYKEKPIQNQAKSVDVEYTVQFTPLNPDDDFRPGLKDTKLLKTLA IGDTITSQELLAQAQSILNKNHPGYTIYERDSSIVTHDNDIFRTILPMDQEFTYRVK 20 NREOAYRINKKSGLNEEINNTDLISEKYYVLKKGEKPYDPFDRSHLKLFTIKYVDVD TNELLKSEQLLTASERNLDFRDLYDPRDKAKLLYNNLDAFGIMDYTLTGKVEDNHDD TNRIITVYMGKRPEGENASYHLAYDKDRYTEEEREVYSYLRYTGTPIPDNPNDK (SEQ ID NO:18)

Table 9: DNA sequence of SK from S. equisimilus H46A

ccaaaactaca tttctagcag tggatttcaa caaacatggt taaggccaaga 101 ctttattgaa gttgcttgtc 201 taggcaaaat gacctcaagc gacataaaaa tgctgtttgg gttgtgctga 201 taggcaaaat gacctcaagc ctgatgatg aagacaataaaa tctgagaagt 201 caaacagc tgagggggat tgccctgatg atcaagcaaa taccgactgc 201 aaggtagacc taggggtgta aagacctcat attgaccaa cccaccta 202 aaggtagacc taggggggat tgccctgatg atcaagcaaa taccggtgcc 203 aaggtagacc tcttttc gacac tagacacaa taccgctgcc 204 ttggccct ctttgacac taccaccta tttgaccaa cccaccta 205 ttcatgacag tctttaaacc attaacaca tttatcct aacgacattca 206 ttcatgaga ttttataatt attattaacac attactatacac 207 ttttttcatg tgcattaaac ttttttagcaa gacattcaa ttaacacacgt 208 tttttttcatg tgcattaaaa tagtattta tctgctttt atcacacagt 209 tttttttcatg gagtttctat tttttagca gaacattcaa ttaatatta 209 sactattag aggtttctat tagaaaaatac attaatttac 209 sol ctggacctga tggctgcta gacgtccat 209 sol ctgacctga tggctgcta gacgtccat 200 sol taaattttt gaacacat ttggaacagt 201 taaattttt gaacacat ttggaacagt 202 sol taaattttt gaacacat ttggaacagt 203 sol taaattttt gaacacat tggtagacga 203 sol tatttgcaag cgttaagcca aaatcaaac 204 sol taaattttt gaacacgtc caatcacaca 205 sol taaattgat gctaacgtcc acagtaacga 206 sol taaattgat gctaacgtcc acagtaacga 207 sol tttgcaag ggatgcaac cacgtaacga 208 sol tttgcaag ggatgcaac cacgtaacga 208 sol tttgcaag ggatgcaac cacgtaacga 209 sol taaattgat gctaacgtc cacagtaacga 209 sol taaattgat gctaacgtc cacagtaacga 200 sol taaattgat gctaacgtc cacagtaacga 200 sol taaattgat gacacgtcac cacgtaacga 201 sol taaacacgag aacacgtcac cacgtaacga 202 sol tattgcaag ggacatgtgc 203 sol tattgcaag ggacatgtgc 203 sol tattgcaag ggacatgtgc 204 sol taaacacacag gaaccgtaccacacacacacacacacacacacacacacac							
cottattgaa gttgottgtc gacataaaaa tgotgttgg gttgtgcga taggcaagac cottcaagcg gttgdgcgaagaggat tagcacaacaacacacacacacacacacacacacacacac			ctgcagctac	ctgataccag	gcatttccaa	caaacatggt	taaggccaaa
101 ctttattgaa gttgcttgtc gacataaaaa tgctgtttgg gttgtgctga taggcaaata gagccaaaca cctgctgcaaaca cctgcaaaca ctgctgggggggggg		51	ccaaaatcac	tttctagcgt	tggcaagaga	ccttcaagcg	agcgcaagac
151 taggcaaaat gacctcaagc ctgcaatca tctgctggag caactcaact			ctttattgaa	gttgcttgtc	gacataaaaa	tactatttaa	attatactaa
agtocacte grammaccty ctgatgatty aggtamatam actgagaagt ctcamacage taggggggat tgccctgatg actacagcama taccgctgcc aggatamacatg tagcggctgc aggatamacatg attgagaaa taccgctgcc aggatamacatg gattatggama attgagaagtg ctccttttic gattagacata taccgccactc tattaccca acccacactc tattaccaca acccacactc tattaccaca acccacactc tattaccaca acccacactc tattaccaca acccacactg tattaccacact tattaccaca acccacactg tattaccacacactg tattaccacacacactg tattaccacacactg tattaccacacacactg tattaccacacactg tattaccacacacactg tattaccacacactg tattaccacacactg tattaccacacactg tattaccacacactg tattaccacacactg tattaccacacactg tattaccacacactg tattaccacacactacacactacacactacacacactac	5		taggcaaaat	gacctcaagc	cctgcaatca	tctgctggag	caactcaact
ctcaaacagc tgaggggat tgccctgatg atcaagcaaa taccgctgcc aggtagacc tagcgggtgac aggacctcat attgacccaa cccacctca 351 agtaataagc gctcttttc ggataaacat gatttgggaa 451 tggtcccct tctttgacac tcaccacctc tttatctcct aacggatgag 451 ggcctacttg catctctgga aaatagtctt ttagctccat agccattcct 551 tcatgacgg tctttaaacc attataacac agaccattcat 551 tcagttgtt gtcagcacga ttttgtattt tctgcctttt taaccacagt 551 tcagttgtt gtcagcacga ttttgtattt tctgctttt taacactataa 661 acttatcgga atattaattt atgtttagc taaaaaaagg attattattca 751 tttgtgtctt taaaaccatt atgttatct aaaaaaagg attattatca 751 tttgtgtctt taaaaccatt tggaaaggt cattcgtc 851 acctgtgttt gcactaacat ttggaacagt cattctgtc 851 ctgaacctga gtggctgcta gaccgtccat ctgtcaacaa gtgttgttgc 851 gtgttagcg ttggctgcta gaccgtccat ctgtcaacaa acctgacattg 851 ctgaacctga gtggctgcta gaccgtccat ctgtcaacaa gtgttgttgc 851 ctgaaccag gtggctgcta gaccgtccat ctgtcaacaa gtgttgttgc 851 cagagcaagg cttaagtcca aaatcaaacg acctgctcat aggacagta 851 cagagcaagg cttaagtcca aaatcaaaca catttgctca caggcaatta 851 cagagcaagg cttaagtcca aaatcaaaca catttgctca tggaagggag 851 li51 acaattgatc gctaacgtcc acagtaacga cgactacttt 861 lttgctaaca gagagcatgcg catactgcg 852 li51 acaattgatc gctaacgtcc acagtaacga cgactacttt 863 lttgcaacaa atggttcggt aaccttgccg gaaccgacaa gggtcattg 854 lttgcaacaa atggacgac atactggc gaaccaacacg gaggtcattg 855 lcgtacacaa gtgacggtg acccaacacg gaggtcattg 856 lttgcaacaa atggacgac atactgac attacaagga aaccaacacacacacacacacacacacacaca			aagtcagctg	gtaaaacctg	ctgatgattg	aggtaaataa	actgagaagt
adgitaccc tagoggotge aagacectaat attgacccaa coccaccta agtaataage gotetittic ggataaacat gattigggaa aatgcactat tiggoccacttictic ggataacacta titactect aacgattgag foll titatgacgt tottiaaacc tacccaccte titatetect aacgattgag tottiaaacca attataaca atgactetit taccacacgt tottiaacaca ggattateca tittiageaa gaacattecaa tacacacgt tittiteati gtotgacacaa attitiageaa gaacattcaa tacacacgt tittiteati tittigtatti attitiageaa gaacattcaa tacacacgt tittiteati taccacacga tittigaaca gaacattcaa tacacacgat tittiteati tacacacat tittiageaa gaacattcaa tacattaa aactitiaga agtiticat gaaaaataa tacattita aaaaaatcaa taggititita actititiggacti gaagtiticat gaaaaattac tatactitiggaacgat totggacctag gagiticatat gaaaaattac tacattitiggaacgat actititiggaa gagiticatat gaaaaattac tacattitiggaacgatagag totggacctag titgetggac tacacacacacacacacacacacacacacacacacac			ctcaaacagc	tgagggggat	tgccctgatg	atcaagcaaa	taccactacc
agtaataagc gctctttttc ggataaacat gatttgggaa aatgacacta 451 tggtccct tctttgacac tcaccactc tttatctct aacggatgag 451 tcattgacgt tcattctgga aaatagtctt ttagctccat agccattcct 551 tcattgacgt tctttaaacc attaatacac atgactcttt atcacacagt 551 tcattttgt gtcagcacga tttttgattt tctgcctttt taatcataa 661 aactaaataa gggttattca tttttagcaa gaacattcaa ttaaatagct 761 tttttcatt 751 tttgtgtctt taaacacatt tattttagcaa gaacattcaa ttaaatagct 861 aacttttaag aggtttctat gaaaaattac taaataggg attgaaactt 801 aacttttaag aggtttctat gaaaaattac ttatctttg ggatttgca 851 acttttagg aggttgacac ttggagacag caattcgtg caagcactat 951 ctggacctga gtggctgcta gacggccat 951 ctggacctga gtggctgcta gacggccat 951 caagacagag cttaactgaa aactacaa acctgaccatt 1001 taaattttt gaaatcgac ttaactacac acctgccaatta 1051 cagagcaagg cttaactga gaaagctgac caattcagag caattcagg 1051 cagagcaag cttaactgaa gaaagctgac ttactaaagg ctattcaaga 1051 cagagcaaga gacattgg gaaagctgac ttactaaagg ctattcaaga 1051 cagagcaaga ggacatcga caactagaac cattgctac tggaaagcaga 1051 cagagcaaga ggacatcga caactagaac cattgctac tggaaagcaga 1051 cagagcaaga ggacatggac aactagaac cattacaaga cactgccaa aggacaatca 1051 aaaaaccaag ggaaattgg ggacatggac accaaccag ggaaaccaataa 1351 aaaaaccaag ggacattgg gggttagaca ataacaaaac 1351 aaaaaccaag gaaattggt gagtggaac ataacaagaa ttaactagaca 1351 cgtaaccaa agaagctgac caaccaacaacaacaacaacaacaacaacaacaacaa			aaggtgaccc	tageggetge	aagacctcat	attgacccaa	ccccacctca
451 ggcctacttg catcottgga aaatagtott ttagctccat agcattcct 501 ttcatgacgg tctttaacc attataacca atgactctt atcacacagt 551 tcagtttgtt gtcagcacga tttttgtatt tctgccttt taatcattaa 601 aactaaataa gggttattca ttttagcaa gaacattcaa ttaaatagct 701 tttttcatt gtcattaaaa tatcatttaa aaaaaatcaa taggttttta 751 tttgtgtct taaaaccatt atgttattct ataaataggg attgaaactt 801 aactttaagg atattaatt atcatttaa aaaaaatcaa taggttttta 701 acttttcatt gcattaaaa tatcatttta aaaaaatcaa taggttttta 702 seppending stiggetgetga gacagtccatt ggaacgattgg 851 actgctgttt gcactaacat ttggaacagt caattctgtc caagctattg 951 gttgttaggg ttggctgata gaccgtccat ctgtcaacaa caggccaatta 951 gttgttaaggg ttggctggtaa taaactaaaa catttggtaa agaattcaa 1001 taaattttt 1001 taaattttt 1001 ggaatgcac ataaacttga gaaagctgac ttactaaagg ctattcaaga 1001 ggaatgcac ataaacttga gaaagctgac ttactaaagg ctattcaaga 1001 acaattgatc gctaacgtcc acagtaacga cactactgtg 1101 gcgatgtcac ataaacttga gaaagctgac ttactaaagg ctattcaaga 1201 atttgcaag ggcatgtgcgt aaccttgccg accaacctg tcaagaatt 1201 atttgcaag ggcatgtgcgt accacaacactg gaaacggcaa ggtctactt 1201 atttgcaag ggcatgtgcgt accacaacacga gaccaacatg 1351 aaaaccaag ggcatgtgcgt accacaacacgacacacacacacacacacacacacacac			agtaataagc	gctctttttc	ggataaacat	gatttgggaa	aatgcacata
forestantly catchedge anatogated the tecquency testing	10		ttggtcccct	tctttgacac	tcacccactc	tttatctcct	aacqqatqaq
tcagtrgtt gtcagcacga ttttgtattt tctgccttt taatcacacagt tcagttrgt gtcagcacga ttttgtattt tctgccttt taatcattaa aactaaataa gggttattca tttttagcaa ttttttcatt gtcattaaaa tattatta taaaaaaggt attatttacc ttttttcatt gtcattaaaa tattatta aaaaaaaggt attatttacc ttttttcatt gtcattaaaa tatgttattca aaaaaaggt attatttacc sol attttagga atattaattt atgttattgc sol ttttttcatt gtcattaaaa tatgttattca aaaaaaggt attatttacc sol attttagga atattaatta tatgttattgc sol ttttttgtgtctt taaaaaccatt atgtattca aaaaaatcaa taggttttta sol aacttttagg aggtttctat gaaaaattac caattctgc ggatgttgcac ctggacctga gtggctgcta gaccgtccat ctgtcaacaa cagcaattag ptfgttaggc gtggctgcta gaccgtccat ctgtcaacaa cagcaattag stfgttaggc gttaagtcca aaaatcaaca acctgctcaa aggaaagg sol taaattttt gaaatcgac caagtaacaga catttcgcc sol taaattttt gaaatcgac caagtaacaga cattcagaca acctacaaga sol tttgcaag ggatgcaacc attactgacc gactacttt gaaacggcaa sol tttgcaag ggacatggc gggttagaacc ataaaagaa aaccaataa sol tttgcaag ggacatggc gggttaggaac ataaaagaa aaccaataac sol taaaaccaag ggaaatctgt gggttagaac atacaagaa aaccaataca sol tttaaaccaag ggacattgg gggttaggaac atacaagaa aaccaataca sol ttaaaccatg atgacgatt cagaccaggt ctcaaagaactgac sol taaaaccaaca gcaattggg ggctagaaca atacaagaa aaccaataca sol taaaaccaaca gcaattggg aaccatcaca atctcaagaa ttcaagaatt sol gaaaaccaaca gcaattggg ggctagaacaca acctacaca atctcaagaa sol taaaccatag acaagatta ctaagacaaga aaccaataca sol taaaaccata gcaattggg aacacaccaca acctacaagaat accaacacacacacacacacacacacacacacacac			ggcctacttg	catctctgga	aaatagtctt	ttagctccat	agccattcct
tcagtttgtt gtcagcacga ttttgtattt tctgcctttt taatcattaa aactaaataa gggttattca tttttagcaa gaacattcaa ttaaatagc ttttttcatt gtcattaaaa tatcatttta 751 ttttttcatt gtcattaaaa tatcatttta 751 ttttttcatt gtcattaaaa tatcatttta 751 ttttttcatt gtcattaaaa tatcatttta 751 ttttttcatt gtcattaaaa tatcatttta 801 aactttagg aggtttcat gaaaaattac tatagttttgc 851 actgctgttt gcactaacat ttggaacagt 951 gtgttagcg gtggctgtag gaccgtccat 951 gtgttagcg ttgctggtac tgttgagggg 1001 taaatttttt gaaatcgatc tacaatcacg 1001 taaatttttt gaaatcgatc tacaatcacg 1101 gcgatgca ataaactgac aaatcaaaac catttgctac tggaagagg 1101 gcgatgtca ataacgtca acagtaacga 125 1151 acaattgatc gcataacgtc acacttagacc atactgacc 126 1201 atttgcaag ggacatgtgc gcgttagacc atacatcacg 1301 tttgctaagc ggacatgtgc gcgttagacc atacatgacc 1301 tttgctaagc ggacatgtgc gcgttagacc atacatgaca 1301 tttgctaagc ggacatgtgc gcgttagacc atacatgaca 1351 aaaaccacac gcaattcggtg acccatcacac acagcaatta 1451 gaaaacacta gctatcggtg acccatcacac acagcatatac 1551 cgtgactcac acatcgtcac tcatgacaat atactgaca 1551 cgtgactcac acatcgtcac tcatgacaat 1551 cgtgactcc caatcgtcac tcatgacaat 1551 cgtgactcc caatcgtcac tcatgacaat 1551 cgtgactcc caatcgtcac tcatgacaat 1551 cgtgactcac agaaattac cgtccttaaa 1551 accattagat caacgattac ctaaccatcac 1551 cgtgactcct caatcgtcac tcatgacaat 1551 cgtgactcct gaaacatac ctaaccatcac 1551 cgtgactcac agaaattac cgtccttaaa 1551 accattctg agaaaatata cgtccttaaa 1551 cgtgactcac acatcgaca ttaaaacacac 1701 cttacacac acatccacac tcatgacaat 1701 ctcatcacac acatccacac tcatgacacac acaggcaaact 1701 ctcatcacac acatccacac ttgaaactgt caccacacaaa 1701 ctcacacac acatccacac ttgaaactgt caccacacaaa 1701 ctcacacac acatccacac ttgaacaccac acaggcaacac 1701 ctcacacac acatccacac ttgaacaccac acaggcaacac 1701 ctcacacac acatccacac ttgaacaccac acaggcaacaccacacacacacacacacac			ttcatgacgg	tctttaaacc	attataacac	atgactcttt	atcacacagt
actaaataa gggttattca tttttagcaa gaacattcaa ttaaatagct attatacgt attatacgc ttttttcatt gtcattaaaa tatcatttta aaaaaaatga attattacc ttttttcatt tttgtgtctt taaaacacatt atgttattca aaaaaatcaa taggttttga aacttttaag aggtttcat gaaaaattac ttatctttg ggatgtttgc cagtcgat ctggacctga gtggctgcat gtgacacacat ttggaacagt caattctgtc caagctattg gtggtgca gttgagggg ggttatcat ggaacgtccat ctgtcaacaa acgactattg ggatgttgc gaacacacacacacacacacacacacacacacacaca			tcagtttgtt	gtcagcacga	ttttqtattt	tctacctttt	taatcattaa
701 tttttcatt gtcattaaaa tatcatttta aaaaaatga attattacc			aactaaataa	gggttattca	tttttagcaa	gaacattcaa	ttaaataget
ttttttcatt ttgtgtctt taaaacatta atgatattct aataaagag attggtagaactt ttggaacatt aggattttaa aggattttaa aggattttaa aggattttaa aggattttaa aggattttaa aggattttaa aggattttaa aggattttaa aggatttaat atgataatta ttggaacatt aggattttg ggatgttgg ggagtgggg gggggtgat ggaggatgggg aggatgaaca actggaa actagtca aggaaactaag actagtca aggaaactaag actagtca aggaaaga actagtgga atgaatcaag actagtgga atgaatcaag actagtggaa atgaatcaag actagtgga atgaactaag actagtgga atgaactaag actagtggaa atgaatcaag actagtggaa atgaatcaag actagtggaa atgaatcaag actagtggaa atgagtggaa atgagtgggaa atgagtgggaac atactggag accaacctg tgaggaaga accagtaattg aggaaaga accagtaatgg aggaacggaa atgaggaaga atgggaagaga atgggaagaga atgggaagaga atgggaagaga atgggaagagaagagaagagaagagaagaagagaagagaagag	15		atttatcgga	atattaattt	atgtttatgc	taaaaaaggt	attatttacc
actittagg aggittetat gaaaaatac ttatettitg ggatgittge actgetgtt geactaacat ttggaacagt caattetgte caagetatig gatgittge gacgatega gacggteeat ctggaacag acgatatagtet 1001 taaattitt gaaategate taacateag acctgeteat aggaaaga 1051 caggacagg cttaagteea aaateaaae cattegetea tgatgaggg acgaateaag acattagtet 1010 gegatgteae ataaaettga gaaagetgae ttactaaagg ctatteaggg 1101 gegatgteae geaacgtee acagtaacaa cgattatetgi 1201 attitgeaag egatgeaaee ataaettgat gaaagetgae ttactaaagg ggtetaetti 1251 getgacaaag atggteeggt acceaacetg gaggetattg 1201 attitgeaag egatgeaee attactgate gaggegaa ggtetaetti 1251 getgacaaag atggteeggt aacettgeeg acceaacetg tecaagaatt 1301 ttigetaage ggaaatetgt gatggaae ataataagaea aaaceaatae 1351 aaaaceaage gaaatetgt gatggaaea ataatagaea gattacteee 1501 aageacaag catttaaaea 1551 egtgaceaea geattetaaae aaaaaceae caagetate 1451 gaaaaacata geattetaaae 1551 egtgaceae caategteee caategteee caategteee teaagaaatta 1601 accaatggat caagagtta ctaagacaat teateagete 1701 etgatecteg agaaatata ettacegtgt taaaaaategg gaacaetgte 1701 etgatectetg agaaatata egteetaaa aaaggggaaa agcegataga 1751 teeetttgat egeagteet tgaaactgt eaceactaea aacegtaea 1861 egaaactaea egaattgeta aaaagtgage agcegategae 1901 acgataceaa aatetegage tttataega egeatetaae 1901 acctacaae aatetegag gataateae egeetaaeae aacegtaeta agetagegaa 1951 gaaaagtaga ggataateae gagaaacaae accegtaeae 1901 acctacaae aatetegag ettttaggta taataegga agetetaaa 2001 atgggeaage gaccegaagg agagaatget accegateae acceateaaa tacgttgatg 1951 gaaaagtaga gaccegaagg agagaatget acceateaae aatetegatg acceateaea accegtataa agetagegaa 2051 taaaaagega acctaacet gataaceca acceateaaa tacgttgata 2051 taaaaacga egaccaaga agaaacaca acceateaae acceateaae 1901 atacaggac acceataaee gagaaaaaaa agcagaaaataa 2201 atacaggac acctatacet gaaaaaaaa agcagaaaaa aaceggeta 2201 eagtteettitteta accttaaet gagaaaaaaa agcagaaaaa accaggeta 2201 eagtteetaa ataacegaa acctataacet gaaaaaaaa agcagaaaa agcagaacaa 12201 eagtteetaa ataacegaa acctataacet gacaaaaaa agcagaaaaa agcagaatte 2251 eggtetaaa ataacegaa acctataacet gaca			ttttttcatt	gtcattaaaa	tatcatttta	aaaaaatcaa	taggtttta
851 actyctytt gcactacat tygaacagt caattctyte caagetatty 951 gtgacetga gtgetgeta gacegteeat cyteaacaa cagecaatta 951 gtgttageg ttgetggtae tygtgagggg acgaateaag acattagtet 1001 taaattttt gaaategate taacateaeg acetgeteat aggaaaga 1051 cagageaagg cttaagteea aaateaaae cattgetea tggaagggal 1051 cagageaagg cttaagteea aaateaaae cattgetea tggaagggg 1101 gegatgteae ataaaettga gaaagetgae ttaetaaagg ctatteaaga 25 1151 acaattgate getaacgtee acagtaaega cgactacttt gagggteatty 1201 attttgeaag egatgeaaee attaetgate gaaaeggeaa ggtetaettt 1301 tttgetaage ggacatgteg gegttagaee ataataaagaa aaaceaatae 1351 aaaaceaage ggacatgteg gegttagaee ataataaagaa aaaceaatae 1351 aaaaceaage gaaatetgtt gatgtggaat ataetgtaea gtttaeteee 1501 aageacaaag catttaaae acaceage tecaaagaat taetgaea 1551 agaaacacta getateggtg acaceateae caggetate 1551 aageacaaag catttaaae aaaaaceaee caggetatae gatttatgaa 1551 acaattggte caateggtea teaaaaateag gatttaeteee 1501 aageacaaag catttaaae aaaaaceaee caggetatee 1501 aageacaaag catttaaae aaaaaceaee gatttattgaa 1551 cgtgacteet caategteae teatgacaat gacattttee 1601 accaatggat caaagattta cttaecetgt 1601 accaatggat caaagattta cgteetaaa agaaaataaa caacactgae 1701 ctgatetetg agaaatatta cgteetaaa aaaaggggaaa agcegtatga 1751 tecetttgat cgcagteaet tgaaacetgt caacaacatg 1851 cgtaaceaa cgaattgeta aaaaaggaga agcettaaae 1851 cgtaaceaa actecagag tttataeega agcettaaae 1851 gaaaagtaga ggataateae gataaceaa accgtatea 1801 tegataceaa cagaattgeta taaaaagtaga agcettaaae 1851 gaaaagtaga ggataateae gatgacacaa accgtatea 1901 actetacaae atecegaag attataeega agctataaet 2001 agggeaage gaceegaagg agaaagaaga agttaeaaet 2011 ataaaggaa acctaacet gataaceca accgtateat 2011 ataaaaggaa cactaacet gataaaeaaa agcagaaata 2011 cagttgettg ctttttteta acctettagt tgaagacet 2011 cagttgettg ctttttteta acctettagt gacaaaaaa agcagcactt 2251 cgtgtetaaa ataategaa ctggtecaace actgacaaaaa agcagcactt 2251 cgtgtetaaa ataategaa ctggteacea attgatgaga ctaacectgaa 2301 tactteacee aaaaacecea cgcteaacet ggcacaaaate tgccaattg			tttgtgtctt	taaaaccatt	atgttattct	aataatgggg	attgaaactt
20 901 ctggacctga gtggctgcta gaccgtcatt cagcattag 951 gttgttagcg gtggctgcta gaccgtcatt ctgtgacaaa gttgttgtagcg ttgctgacgtgta tgttgagggg acgaatcaag acattagtct 1001 taaattttt gaaatcgatc taacatcacg 1051 cagagcaaagg cttaagtcca aaatcaaaaa catttgctaa tgtgaaggacgaagga 1051 acaattgatc gctaacgtcc acagtaacga cttatcaaagg ctattcaaagg 25 1151 acaattgatc gctaacgtcc acagtaacga cgactacttt 1251 gctgacaaag atggtteggt aaccttgceg accaacctg gaaacctgt 1301 tttgctaagc ggatgcaacc attactgatc gaaaccagga 1351 aaaaccaag gaaatctgtt gggtagaac atactgaca 1351 aaaaccaag gaaatctgtt gggtagaac atactgaca 1351 aaaaccaag gaaatctgtt gggtagaac atactgaca 1351 aaaaccacag gaaatctgtt gggtagaac atactgtaca 1451 gaaaaccata gctatcggtg gcgttagacc atactgaaa aaaccaatac 1501 aagcacaaag cattttaaac aaaaaccacc ctaaaggat 1451 gaaaacacta gctatcggtg acccatcaca 1501 aagcacaaag cattttaaac aaaaaccacc caggctatac 1501 aagcacaaag cattttaaac aaaaaccacc caggctatac 1501 accaatggat taaaaatcgg catttaaac 1551 cgtgactcct caatggtac tcatgacaat 1551 cgtgactcct caatggtac tcatgacaat 1701 ctgatcctg agaaatatta cgtccttaaa aaaggggaa agcagtttt 1701 ctgatcctg agaaatatta cgtccttaaa aaaggggaa agcagtttt 1801 tccatttgat 1801 tcgataccaa 1851 cgtaactag 1951 gaaaagtaga 2001 atgggcaagc accaacac 2001 atgggcaac 2001 atgggcaac 2001 atgggcaac 2001 atgggcaac 2001 atgggcaac 2001 atgggcaac 2001 ataccacac 2001 cagtgctc 2001 atgggcaac 2001 atgggcaaca 2001 atgggcaaca 2001 atggacaca 2001 atgggcaaca 2001 atggacaca 2001 atggacaca 2001 atggacaca 2001 atggacacacacacacacacacacacacacacacacacac			aacttttagg	aggtttctat	gaaaaattac	ttatcttttg	ggatgtttgc
901 ctggacctga gtggctgcta gaccgtccat ctgtcaacaa cagccaatta 951 gttgttagcg ttgctggtac tgttgagggg accattacaag acattagtct 1001 taaattttt gaaatcgatc taacatcacg accattagtgca 1051 cagagcaagg cttaagtcca aaatcaaaac catttgctac tgatagtgg 1101 gcgatgtcac ataaacttga gaaagctgac ctattcaaga gctattcaaga gctaacgtc acagtaacga cgactacttt 1201 attttgcaag ggtagcaacca attactgatc gaaacgggaa ggtctacttt 1251 gctgacaaag atggttcggt aaccttgccg accaacctg 1301 tttgctaagc ggacatgtgc gcgttagacc atatacagaa atggttcggt aaaccaaccag gaaatcgtt 1451 gaaaaccacag gaaatctgtt gggtggaat atactgtaca aaaaccacac 1501 aagcacaaag catttaaac aaaaaccac caggctatac 1501 aagcacaaag catttaaac aaaaaccac caggctatac 1501 accaatggat caacggtta ctaacagaat 1551 cgtgactcc caatcgtcac tcatgacaat 1501 accaatggat caacgagtta ctacggtg aagaaataa 1701 ctgatcctcg agaaatata cgtccttaaa aaggggaaa 1751 tccctttgat cgagtcact tgaaactgt caacaactgac 1701 ctgatcctag agaatatta cgtccttaaa aaggggaaa 1851 cgtaactag cgaattgcta aaaagtgagc accaaccaa agccgtatga 1951 gaaaagtaga cgaattgcta aaaagtgagc accaaccaa aacctggca 1951 gaaaagtaga 2051 taaaagacg caaccaaca aacctcgaga aaccttaacaa accttaacga 2001 atgggcaagc accaacaga agagaaataca 2001 atgggcaagc accaacaga agagaaataca 2001 atacagggac accaacaga agagaaataca 2001 atgggcaagc accaacagaa agaacagaa agaacagaa agaacagaa agaacagaa 2201 cagttgctg caacagaa accttaacct gaaaaccta accgttaaca 2201 cagttgctg caacagaa agaaacacaa agaaacaaca 2201 cagttgctg caacagaa agaaacacaa agaaacaaca 2201 cagttgctg caacagaa agaaacaaa agaaacaaa accacttaaca 2201 cagttgctg caacagaa agaaacaaa agaaacaaa agaaacaacaa 2201 cagttgctg caacagaa agaaacaaa agaaacaaa agaaacaacaa 2201 cagttgctg caacaacac 2201			actgctgttt	gcactaacat	ttggaacagt	caattctqtc	caagetattg
1001 taaatttttt gaaatcgatc taacatcacg acctgctcat aggaaaga cttaagtcca aaatcaaaac catttgctac tgatagtggc taacgtcc acaaacgtacgacgacgatttgatctgatc	20		ctggacctga	gtggctgcta	gaccgtccat	ctgtcaacaa	cagccaatta
1051 cagagcaagg cttaagtcca aaatcaaaac catttgctac tgatagtggc 1101 gegatgtcac ataaacttga gaaagctgac ttactaaagg ctattcaaga 251151 acaattgatc gctaacgtcc acagtaacga gaaacggcaagtgacacttt 1251 gctgacaaag atggttcggt acccaacctg gagtcaactt 1251 gctgacaaag atggttcggt aaccttgaca accaacctg gaaaccaatca 1351 aaaaccaagc gaaatctgtt ggttagacc atataaagaa aaaccaatac 1351 aaaaccaag gaaatctgtt acgacaagtt 1451 gaaaacacta gctatcggtg acaccatcac atctcaagaa ttactagctc 1501 aagcacaaag catttaaac acaccacacacacacacacacacacacac			gttgttagcg	ttgctggtac	tgttgagggg	acgaatcaag	acattagtct
1051 cagagcaagg cttaagtcca aaatcaaaac catttgctac tgatagtggc 1101 gegatgtcac ataaacttga gaaagctgac ttactaaagg ctattcaaga 251151 acaattgatc gctaacgtcc acagtaacga gaaacggcaagtgacacttt 1251 gctgacaaag atggttcggt acccaacctg gagtcaactt 1251 gctgacaaag atggttcggt aaccttgaca accaacctg gaaaccaatca 1351 aaaaccaagc gaaatctgtt ggttagacc atataaagaa aaaccaatac 1351 aaaaccaag gaaatctgtt acgacaagtt 1451 gaaaacacta gctatcggtg acaccatcac atctcaagaa ttactagctc 1501 aagcacaaag catttaaac acaccacacacacacacacacacacacac		1001	taaattttt	gaaatcgatc	taacatcacg	acctgctcat	aggaaaga
1101 gcgatgtcac ataaacttga gaaagctgac ttactaaagg ctattcaaga 1201 attttgaaag cgatgaacc acagtaacga gaaacggcaa gggtctacttt 1251 gctgacaaag atggtcagt gaaccttgcc gaaacggcaa gggtctacttt 1301 tttgctaagc ggaaatgtgc gcgttagacc accaacggca atactggca acccaacctg tccaaagata 1301 tttgctaagc ggaaatgtgc gcgttagacc atataaagaa aaaccaatac 1351 aaaaccaagc gaaatctgtt gagtggaat atactgtaca gtttactccc 1501 aagcacaaag catttaaac acacacaca atccaaagaa ttactaggtc 1501 aagcacaaag catttaaac aaaaaccacc caggctatac gatttatggaa 1551 cgtgactcct caatcgtcac tcatgacaat 1601 accaatggat caaaaaatct ggtctgaatg agcattttcc gaacaagctt 1701 ctgatcctg agaaatatta cgtccttaaa aaagagggaaa agccgtatga 1751 tccctttgat cgcagtcact tgaaactgt caccatcaaa tacgttgatga 1851 cgtaactaa actcagaag atttaacgac attacagaa 1851 cgtaactaa actcagaag atttaacgac acttaacgaga 1851 cgtaactaa actccagaga atttaacgat 2001 accaatcgac gaccagaag ggataatcac gaatgacacca accgtatcat aaccttaactg 1951 gaaaagtaga ggataatcac gatgacacca accgtatcat accttaactg 1951 gaaaagtaga ggataatcac gatgacacca accgtatcat taggctaaact 2001 atcacaac accataacc gatgacacca accgtatcat taggctatcat 2001 atacagggac acctatacct gatgacacca accgtatcat accttaactg 2051 taaaagatcgt cttttttcta acctcttagt 2251 cgtgctaaaa ataatcgtaa ctggtcaatca 2201 cagttgctc aaaaacgcca cgctcaactg 2351 tcattaaagc gatcataaaa ctggctagcc aaaaacccta 2301 tactgccc aaaaacccc aaaaacccc ctttttttca acctcttagt tgaagacat tgccaattgt 2351 tcattaaagc gatcataaaa ctggctagcc ataaccgca 2301 tactgccc aaaaacccc ctttttttca acctcttagt tgcagctcct tgcagctcct 2351 tcattaaagc gatcataaaa ctggctagcc ctaatcacc gatcaaaaa accacatcgca 2401 gtaaaggctg ggcgattcc cttttttttgtgtg tcagcaaaaac tgccaacta ggcaaaata ggcagatatgc		1051	cagagcaagg	cttaagtcca	aaatcaaaac	catttoctac	tgatagtggc
151 acaattgate getaacgtee acagtaacga egactacttt gaggteattg 1201 attttgcaag egatgcaace attactgate gaaacggcaa ggtetacttt 1251 getgacaaag atggteeggt aacettgeeg acceaacetg tecaagaatt 1301 tttgetaage ggacatgtge gegttagace ataataaagaa aaaccaatae 1351 aaaaccaag gaaatetgtt gatgtggaat atactgtaca gtttacteec 1451 gaaaacacta getateggtg acaccateac atectaaagaa tactgatea 1501 aagcacaaag catttaaac aaaaacace caggetatae gattatggaa 1551 cgtgacteet caategteac teatgacaat 1551 cgtgacteet caategteac ataacgteac gattattggaa 1551 cgtgacteet caategteac teatgacaat gacattttee gaagagttt 1601 accaatggat caatgteac teatgacaat gacatttee gaagagttt 1701 ctgatecteg agaaatatta cgteettaaa aagagggaaa agcegtatga 1701 tegatecea egaattgeta agaaatata cgteettaaa aagggggaaa agcegtatga 1801 tegataccaa cgaattgeta aaaagtggg agcegteate 1951 gaaaagtaga ggataateac gaattacagag agaaatget taaggecagag aggaatget taaggecagag aggaatget accttaacagag acctttaac accttaacegag agagaatget agctateat taggectaaact 1951 teataaacga gataaceca accgtacet tgaaacgga aggatateat taggetaaact 2001 atgggecage tataccgaag agagaatget agctateat tagcetatga 2051 taaagategt caccaacaca accgtacet tgaaaccga aggaaatget agctateat accateggtet 2151 tetaaaacga tagagataac cggtaaaceca accategac accaacact accategac accaacact accategac accaacact accategac accaacact accategac accaacact accacegace accaacact accacegace accaacact accacegace accaacact accacegace accaacact accacegace accaacact accacegace accacactaacact accacegace accacactaacact accacegacacacacaccacacacacacacacacacacac		1101	gcgatgtcac	ataaacttga	gaaagctgac	ttactaaagg	ctattcaaga
1201 attitigaag cgatgcaace attactgate gaaacggcaa ggtctactti 1251 gctgacaaag atggtteggt aaccttgeeg acceaacctg tccaagaatt acceaacactg tttacteec gaaatctgtt ggacattgte gaaacacta gctatcece gaagettt cagacaaggt ctcaaagaat atactgtaca caggetatac gatttatgaa gacatttte gaacattt to aagaatatta cagaggtta accatcaca atactgacaga agacatttte gaacatgac atactgacaga agacatacac aagagggaa agcgtatga agacatacac aagagggaa agcgtataca aggacacaca accatcaca accatcaca atactgacga agccgtatga accatcaca accatcaca accgtacaca accgtacacaca accgtacacacacacacacacacacacacacacacacaca	25	1151	acaattgatc	gctaacgtcc	acaqtaacqa	cgactacttt	gaggtcattg
1251 gctgacaaag atggtteggt aaccttgeeg accaacctg tecaagaatt attagee gaaatetgtt ggtatagac atataaagaa aaaccaatac atatacaagaa tatataagaa aaaccaatac atatacaagaa atactgtee gaaatetgtt gatgtgaat tatactgtee 1501 gaaaacacta gctateggtg acaccateac caaggetatac etaaggetat tatacagaa aaaccaatac atactgtee 1501 accaatggat caatetgteac teatgacaat gacatttee gattateggt acaccateac atactgaaga ttactgaaga ttactgaaga ttactgaaga ttactgaaga ttactgaaga ttactgaaga ttactgaagactt accaatggat caatetgeeg acaccateac caaggetatac gattatagaa gacatttee gaacaatttee gacaatttee gacaaataa accaetgaa aggacaataa tacgttgatg agetettaac aggetaaact ettttggtat tatacgae accttaacet gacaaaata accgtateat accttaactga aggaaatge acctatacat accgtateat accttaactga aggacaacca aggacaaata accaetggte agaaaccaa accgtatea accttaactga aggacaacca accgtateat accttaactga aggacaacaa accgtateat accttaactga aggacaccaa accgtateat accttaactga acctataca accttaacct acctetagt tgacaaaaaa accaactgac acctatacat accgtateat accttaactga acctatacat acctetagt tgacaaaaaa accaactgac acctatacat acctetaga acctatacat acctetagac acctatacat acctetagac acctatacat acctetaga acctatacat acctetagac acctatacat a		1201	attttgcaag	cgatgcaacc	attactgatc	gaaacggcaa	ggtctacttt
1351 aaaaccaagc gaaatctgtt gatgtggaat atactgtaca gtttactccc atgaccaggt ctcaaagata ctaagctatt tascaccaccaccaccaccaccaccaccaccaccaccacca		1251	gctgacaaag	atggttcggt	aaccttgccg	acccaacctq	tccaagaatt
1401 ttaaaccctg atgacgattt cagaccaggt ctcaaagata ctaagctatt 1451 gaaaacacta gctatcggtg acaccatcac 1501 aagcacaaag cattttaaac aaaaaccacc caggctatac gatttatgaa 1551 cgtgactcct caatcgtcac tcatgacaat 1601 accaatggat caacaatgtac caatcgtcac 1701 ctgatctctg agaaatatta 1751 tccctttgat 1751 tccctttgat 1751 tccctttgat 1751 tccctttgat 1751 tccctttgat 1751 cgtaaccaa 1751 tccctttgat 1751 tcccttacac 1751 tccc		1301	tttgctaagc	ggacatgtgc	gcgttagacc	atataaagaa	aaaccaatac
1451 gaaaacacta gctatcggtg acaccatcac atctcaagaa ttactagctc 1501 aagcacaaag cattttaaac aaaaaccacc caggctatac gatttatgaa 1551 cgtgactcct caatcgtcac tcatgacaat gacattttcc gtacgatttt 1601 accaatggat caagagttta cttaccgtgt taaaaatcgg gaacaagctt 1651 ataggatcaa taaaaaatct ggtctgaatg aagaaataaa caacactgac 1701 ctgatctctg agaaatatta cgtccttaaa aaaggggaaa agccgtatga 1751 tccctttgat cgcagtcact tgaaactgtt caccatcaaa tacgttgatg 1801 tcgataccaa cgaattgcta aaaagtgagc agctcttaac agctagcgaa 1851 cgtaacttag acttcagagg tttatacgat cctcgtgata aggctaaact 1901 actctacaac aatctcgatg gataaccac accgtatcat accttaactg 1951 gaaaagtaga ggataatcac gatgacacca aggaaatgac 2051 taaagatcgt tataccgaag agaaatgac accgaaata accacggtct 2101 atacagggac acctatacct gataacccta aggaaatgac 2201 cagttgctg tgagattaac tgacaaaaaa agcaagcaac atgctatcaa 2201 cagttgcttg ctttttcta acctctagt tgtagagact tatctgccc 2351 tcattaaagc gatcataaaa ctggctagcc ctttttggtgt tcagcataaa tgcaaattggtaaattggtaaaggctg gatcataaaa ctggctagcc atatcagct tgcaaattggt 2401 gtaaaggctg ggcgatttcc ctttttggtgt tcagcataaa gggtaaattg		1351	aaaaccaagc	gaaatctgtt	gatgtggaat	atactgtaca	gtttactccc
1501 aagcacaaag cattttaaac aaaaaccacc caggctatac gatttatgaa 1551 cgtgactcct caatcgtcac tcatgacaat gacattttcc gtacgatttt 1601 accaatggat caaggttta cttaccgtgt taaaaaatcg gaacaaggct 1701 ctgatctctg agaaatatta cgtccttaaa aaaggggaaa 1751 tccctttgat cgcagtcact tgaaactgtt caccatcaaa aacggggaaa 1751 tccctttgat cgcagtcact tgaaactgtt caccatcaaa tacgttgatg 1801 tcgataccaa cgaattgcta aaaaggggaa agctcttaac agctagcgaa 1851 cgtaacttag acttcagaga tttatacgat cctcgtgata aggctaaact 1901 actctacaac aatctcgatg gataatcac gatgacacca aggaaatgct aacgtatcat aacgttaat 2001 atgggcaagc gacccgaagg aggaaatgct aacgtatcat aagacagta 2051 taaagatcgt tataccgaag accataacc gataaccct aagacaataa accacggttat 2101 atacagggac acctatacct gataacccta aggaaaacaa agcaagcaaata accacggtt 2251 cgtgtctaaa cctttttcta acctcttagt tgtagagact atgcatacaa 2201 cagttgcttg ctttttcta acctcttagt tgtagagact atgcatacaa 2301 tatctgccc aaaaacgcca cgctcaactg gacaaaata tgcaaattgt 2351 tcataaagc gatcataaaa ctggctagcc tttttggtg tcagaattgt gacaaaatc tgcaaattgt 2401 gtaaaggctg ggcgatttcc ctttttggtg tcagcaaaat tgcaaattgt	30	1401	ttaaaccctg	atgacgattt	cagaccaggt	ctcaaagata	ctaagctatt
1551 cgtgactcct caatcgtcac tcatgacaat gacattttcc gtacgatttt 1601 accaatggat caagagttta cttaccgtgt taaaaaatcgg gaacaagctt gaacaagctt 1651 ataggatcaa taaaaaatct ggtctgaatg aagaaataaa caacactgac 1701 ctgatctctg agaaatatta cgtccttaaa aaaggggaaa agccgtatga 1751 tccctttgat cgcagtcact tgaaactgtt caccatcaaa tacgttgatg 1801 tcgataccaa cgaattgcta aaaaggtgagc agctcttaac agctagcgaa 1851 cgtaacttag acttcagaga tttatacgat cctcgtgata aggctagcgaa 1851 gaaaagtaga ggataatcac gatgacacca accgtatcat accttaactg 1951 gaaaagtaga ggataatcac gatgacacca accgtatcat accttaactg 1951 gaaaagtaga ggataatcac gatgacacca agcgatacat taggctatat 2001 atgggcaagc gacccgaagg agagaatgct agctatcatt taggctatga 2051 taaagatcgt tataccgaag agaaacgaga agtttacagc tacctgcgtt 2101 atacagggac acctatacct tagacaaaaa agcaagcaac atgctatcaa 2201 cagttgcttg ctttttcta acctcttagt tgtagagact agtgacattt 2251 cgtgtctaaa ataatcgtaa ctggccatc attgatgaga ctaacctgca 2301 tatctgccc aaaaacgcca cgctcaactg gacaaaaatc tgccaattgt 2351 tcattaaagc gatcataaaa ctggctagcc atatcagct tgcagctcct 2401 gtaaaggctg ggcgatttcc ctttttggtg tcagcataaa gggtaaattg		1451	gaaaacacta	gctatcggtg	acaccatcac	atctcaagaa	ttactagete
1601 accaatggat caagagttta cttaccgtgt taaaaatcgg gaacaagctt 1651 ataggatcaa taaaaaatct ggtctgaatg aagaaataaa caacactgac 1701 ctgatctctg agaaatatta cgtccttaaa aaagggggaaa agccgtatga 1751 tccctttgat cgcagtcact tgaaactgtt caccatcaaa tacgttgatg 1801 tcgataccaa cgaattgcta aaaagtgagc agctcttaac agctagcgaa 1851 cgtaacttag acttcagaga tttatacgat cctcgtgata aggctaaact 1901 actctacaac aatctcgatg gataatcac gatgacacca accgtatcat accgttaat 2001 atgggcaagc gacccgaagg agagaatgct aggtatcat agcctatga 2051 taaagatcgt tataccgaag agagaatgct aggtatcat accacggttt 2101 atacagggac acctatacct gataacccta aggaaacgaa agtttacagc tacctggtt 2251 tctaaaacga tgagattaac ctggtcaat agcaagaac atgctatcaa 2201 cagttgcttg ctttttcta acctcttagt tgtagagact atgctatcaa 2301 tatctgccc aaaaacgcca cgctcaactg gcacaaaatc tgccaattgt 2351 tcattaaagc gatcataaaa ctggctagcc atatcagct tgcagctcct 2401 gtaaaggctg ggcgatttcc ctttttggtg tcagcataaa gggtaaattg		1501	aagcacaaag	cattttaaac	aaaaaccacc	caggctatac	gatttatgaa
1701 ctgatctctg agaaatatta cgtccttaaa aaaggggaaa agccgtatga 1751 tccctttgat cgcagtcact tgaaactgtt caccatcaaa tacgttgatg 1801 tcgataccaa cgaattgcta aaaagtgagc agctcttaac agctagcgaa 1851 cgtaacttag acttcagaga tttatacgat cctcgtgata aggctaaact 1901 actctacaac aatctcgatg cttttggtat tatggactat accttaactg 1951 gaaaagtaga ggataatcac gatgacacca accgtatcat accttaactg 1951 gaaaagtaga ggataatcac gatgacacca accgtatcat tagcctatga 2001 atgggcaagc gacccgaagg aggaaatgct agctatcat tagcctatga 2051 taaagatcgt tataccgaag aggaaatgct acctacact gataacccta 2151 tctaaaacga tgagattaac tgacaaaaaa agcaagcaac atgctatcaa 2201 cagttgcttg cttttttcta acctcttagt tgtagagact agtgacattt 2251 cgtgtctaaa ataatcgtaa ctggtccatc attgatgaga ctaacctgca 2301 tatctgcccc aaaaacgcca cgctcaactg gcacaaaatc tgccaattgt 2351 tcattaaagc gatcataaaa ctggctagcc atatcagct tgcagctcct 2401 gtaaaggctg ggcgatttcc ctttttggtg tcagcataaa		1551	cgtgactcct	caatcgtcac	tcatgacaat	gacattttcc	gtacgatttt
1701 ctgatetetg agaaatatta cgteettaaa aaaggggaaa ageegtatga 1751 teeetttgat egeagteaet tgaaactgtt eaceateaaa taegttgatg 1801 tegataceaa egaattgeta aaaagtgage agetettaae agetagegaa 1851 egtaacttag actteagaga tttataegat eetegtgata aggetaaeet 1901 actetacaae aatetegatg ettttggtat tatggaetat aeettaaetg 1951 gaaaagtaga ggataateae gatgaeaeea aeegtateat ageetateat 2001 atgggeaage gaeeegaagg agaaatget agetateat tageetatga 2051 taaagategt tataeegaag agaaaeeeta aeegtateat 2151 tetaaaaega eeetataeet gataaeeeta ageaageaae atgetateaa 2201 eagttgettg ettttteta aeetettagt etgagaeet 2151 egtgtetaaa ataategtaa eeetettagt etgagagaet atgetateaa 2301 tatetgeeee aaaaaegeea egeteaaetg geaeaaaate tgeeaattgt 2351 teattaaage gateataaaa etggetagee atateaget tgeageteet 2401 gtaaaggetg ggegatttee etttttggtg teageataaa gggtaaattg		1601	accaatggat	caagagttta	cttaccgtgt	taaaaatcgg	gaacaagctt
1751 tecettigat egeagteact tgaaactgtt caccatcaaa taegitgatg 1801 tegataccaa egaattgeta aaaagtgage agetettaac agetagegaa 1851 egtaacttag actteagaga tttataegat eeteegtgata aggetaaact 40 1901 actetacaac aatetegatg etttiggtat tatggaetat acettaaetg 1951 gaaaagtaga ggataateac gatgacacca acegtateat acettaaetg 2001 atgggeaage gaceegaagg aggaatget aggatateat taggeetatga 2051 taaagategt tataeegaag aagaaegaga agtttaeage 2101 ataeagggac acetataeet gataaeeceta aegacaaata aecaeggtet 2101 ataeagggac etttitteta aectettagt tgtagagaet atgetateaa 2201 cagttgettg etttitteta aectettagt tgtagagaet atgetateaa 2201 cagttgettg etttitteta aectettagt tgtagagaet etaaeetgea 2301 tatetgeeee aaaaaegeea egeteaaetg geacaaaate tgeeaattgt 2351 teataaage gateataaaa etggetagee atateaget tgeageteet 50 2401 gtaaaggetg ggegatttee ettittggtg teageataaa ggggtaaattg	35	1651	ataggatcaa	taaaaaatct	ggtctgaatg	aagaaataaa	caacactgac
1801 tcgataccaa cgaattgcta aaaagtgagc agctcttaac agctagcgaa 1851 cgtaacttag acttcagaga tttatacgat cctcgtgata aggctaaact 40 1901 actctacaac aatctcgatg cttttggtat tatggactat accttaactg 1951 gaaaagtaga ggataatcac gatgacacca accgtatcat accgttata 2001 atgggcaagc gacccgaagg agagaatgct agctatcatt taggctatga 2051 taaagatcgt tataccgaag aagaacgaga agtttacagc tacctgcgtt 2101 atacagggac acctatacct gataacccta acgacaaata accacggtct 2151 tctaaaacga tgagattaac tgacaaaaaa agcaagcaac atgctatcaa 2201 cagttgcttg ctttttcta acctcttagt tgtagagact agtgacattt 2251 cgtgtctaaa ataatcgtaa ctggtccatc attgatgaga ctaacctgca 2301 tatctgcccc aaaaacgcca cgctcaactg gcacaaaatc tgccaattgt 2351 tcattaaagc gatcataaaa ctggctagcc atatcagct tgcagctcct 2401 gtaaaggctg ggcgatttcc ctttttggtg tcagcataaa gggtaaattg		1701	ctgatctctg	agaaatatta	cgtccttaaa	aaaggggaaa	agccgtatga
1901 actctacac aatctcgatg cttttggtat tatggactat accttaactg 1951 gaaaagtaga ggataatcac gatgacacca accgtatcat accgttat 2001 atgggcaagc gacccgaagg agagaatgct agctatcatt taggctatga 2051 taaagatcgt tataccgaag agagacgaga agtttacagc tacctgcgtt 2101 atacagggac acctatacct gataacccta acgacaaata accacggtct 2151 tctaaaacga tgagattaac tgacaaaaaa agcaagcaac atgctatcaa 2201 cagttgcttg ctttttcta acctcttagt tgtagagact agtgacattt 2251 cgtgtctaaa ataatcgtaa ctggtccatc attgatgaga ctaacctgca 2301 tatctgcccc aaaaacgcca cgctcaactg gcacaaaatc tgccaattgt 2351 tcattaaagc gatcataaaa ctggctagcc atatcagct tgcagctcct 2401 gtaaaggctg ggcgatttcc ctttttggtg tcagcataaa ggggtaaattg		1/51	tccctttgat	cgcagtcact	tgaaactgtt	caccatcaaa	tacgttgatg
1901 actctacaac aatctcgatg cttttggtat tatggactat accttaactg 1951 gaaaagtaga ggataatcac gatgacacca accgtatcat aaccgtttat 2001 atgggcaagc gacccgaagg agagaatgct agctatcatt tagcctatga 2051 taaagatcgt tataccgaag aagaacgaga agtttacagc tacctgcgtt 2101 atacagggac acctatacct gataacccta acgacaaata accacggtct 45 2151 tctaaaacga tgagattaac tgacaaaaaa agcaagcaac atgctatcaa 2201 cagttgcttg ctttttcta acctcttagt tgtagagact agtgacattt 2251 cgtgtctaaa ataatcgtaa ctggtccatc attgatgaga ctaacctgca 2301 tatctgcccc aaaaacgcca cgctcaactg gcacaaaatc tgccaattgt 2351 tcattaaagc gatcataaaa ctggctagcc atatcagctt tgcagctcct 50 2401 gtaaaggctg ggcgatttcc ctttttggtg tcagcataaa gggtaaattg		1801	tcgataccaa	cgaattgcta	aaaagtgagc	agctcttaac	agctagcgaa
1951 gaaaagtaga ggataatcac gatgacacca accgtatcat aaccgtttat 2001 atgggcaagc gacccgaagg agagaatgct agctatcatt tagcctatga 2051 taaagatcgt tataccgaag aagaacgaga agtttacagc tacctgcgtt 2101 atacagggac acctatacct gataacccta acgacaaata accacggtct 45 2151 tctaaaacga tgagattaac tgacaaaaaa agcaagcaac atgctatcaa 2201 cagttgcttg ctttttcta acctcttagt tgtagagact agtgacattt 2251 cgtgtctaaa ataatcgtaa ctggtccatc attgatgaga ctaacctgca 2301 tatctgccc aaaaacgcca cgctcaactg gcacaaaatc tgccaattgt 2351 tcattaaagc gatcataaaa ctggctagcc atatcagctt tgcagctcct 50 2401 gtaaaggctg ggcgatttcc ctttttggtg tcagcataaa gggtaaattg	40	TROT	cgtaacttag	acttcagaga	tttatacgat	cctcgtgata	aggctaaact
2001 atgggcaage gaccegaagg agagaatget agetateatt tageetatga 2051 taaagategt tatacegaag aagaacgaga agtttacage tacetgegtt 2101 atacagggac acctatacet gataacecta acgacaaata accaeggtet 45 2151 tetaaaacga tgagattaac tgacaaaaa agcaagcaac atgetateaa 2201 cagttgettg ettttteta acctettagt tgtagagaet agtgacattt 2251 egtgtetaaa ataategtaa etggtecate attgatgaga etaacetgea 2301 tatetgeee aaaaacgeea egeteaaetg geacaaaate tgecaattgt 2351 teattaaage gateataaaa etggetagee atateagett tgeageteet 50 2401 gtaaaggetg ggegatttee etttttggtg teagcataaa gggtaaattg	40	1901	actctacaac	aatctcgatg	cttttggtat	tatggactat	accttaactg
2051 taaagatcgt tataccgaag aagaacgaga agtttacagc tacctgcgtt 2101 atacagggac acctatacct gataacccta acgacaaata accacggtct 2151 tctaaaacga tgagattaac tgacaaaaa agcaagcaac atgctatcaa 2201 cagttgcttg cttttttcta acctcttagt tgtagagact agtgacattt 2251 cgtgtctaaa ataatcgtaa ctggtccatc attgatgaga ctaacctgca 2301 tatctgccc aaaaacgcca cgctcaactg gcacaaaatc tgccaattgt 2351 tcattaaagc gatcataaaa ctggctagcc atatcagctt tgcagctcct 50 2401 gtaaaggctg ggcgatttcc ctttttggtg tcagcataaa gggtaaattg		1921	gaaaagtaga	ggataatcac	gatgacacca	accgtatcat	aaccgtttat
2101 atacagggac acctatacct gataacccta acgacaaata accacggtct 2151 tctaaaacga tgagattaac tgacaaaaaa agcaagcaac atgctatcaa 2201 cagttgcttg cttttttcta acctcttagt tgtagagact agtgacattt 2251 cgtgtctaaa ataatcgtaa ctggtccatc attgatgaga ctaacctgca 2301 tatctgccc aaaaacgcca cgctcaactg gcacaaaatc tgccaattgt 2351 tcattaaagc gatcataaaa ctggctagcc atatcagctt tgcagctcct 50 2401 gtaaaggctg ggcgatttcc ctttttggtg tcagcataaa gggtaaattg	45	2001	atgggcaage	gacccgaagg	agagaatgct	agctatcatt	tagcctatga
45 2151 totaaaacga tgagattaac tgacaaaaaa agcaagcaac atgctatcaa 2201 cagttgcttg cttttttcta acctcttagt tgtagagact agtgacattt 2251 cgtgtctaaa ataatcgtaa ctggtccatc attgatgaga ctaacctgca 2301 tatctgccc aaaaacgcca cgctcaactg gcacaaaatc tgccaattgt 2351 tcattaaagc gatcataaaa ctggctagcc atatcagctt tgcagctcct 50 2401 gtaaaggctg ggcgatttcc ctttttggtg tcagcataaa gggtaaattg		2051	taaagatcgt	tataccgaag	aagaacgaga	agtttacagc	tacctgcgtt
2201 cagttgcttg cttttttcta acctcttagt tgtagagact agtgacattt 2251 cgtgtctaaa ataatcgtaa ctggtccatc attgatgaga ctaacctgca 2301 tatctgcccc aaaaacgcca cgctcaactg gcacaaaatc tgccaattgt 2351 tcattaaagc gatcataaaa ctggctagcc atatcagctt tgcagctcct 50 2401 gtaaaggctg ggcgatttcc ctttttggtg tcagcataaa gggtaaattg		2101	atacagggac	acctatacct	gataacccta	acgacaaata	accacggtct
2251 cgtgtctaaa ataatcgtaa ctggtccatc attgatgaga ctaacctgca 2301 tatctgcccc aaaaacgcca cgctcaactg gcacaaaatc tgccaattgt 2351 tcattaaagc gatcataaaa ctggctagcc atatcagctt tgcagctcct 50 2401 gtaaaggctg ggcgatttcc ctttttggtg tcagcataaa gggtaaattg		2151	tctaaaacga	tgagattaac	tgacaaaaaa	agcaagcaac	atgctatcaa
2301 tatctgcccc aaaaacgcca cgctcaactg gcacaaaatc tgccaattgt 2351 tcattaaagc gatcataaaa ctggctagcc atatcagctt tgcagctcct 50 2401 gtaaaggctg ggcgatttcc ctttttggtg tcagcataaa gggtaaattg		2201	cagttgcttg	cttttttcta	acctcttagt	tgtagagact	agtgacattt
2351 tcattaaagc gatcataaaa ctggctagcc atatcagctt tgcagctcct 50 2401 gtaaaggctg ggcgatttcc ctttttggtg tcagcataaa gggtaaattg		2251	cgtgtctaaa	ataatcgtaa	ctggtccatc	attgatgaga	ctaacctgca
50 2401 gtaaaggetg ggegatttee etttttggtg teageataaa gggtaaattg		2301	tatctgcccc	aaaaacgcca	cgctcaactg	gcacaaaatc	tgccaattgt
2451 cgacacagat aagatactac cettgatgte teagcataaa gggtaaattg		235I	tcattaaagc	gatcataaaa	ctggctagcc	atatcagett	tgcagctcct
2451 Cyacacagat aagatactac cettgatgte ttggatagae tgatteatet	50	24UI	graaaggerg	ggcgatttcc	ctttttggtg	tcagcataaa	gggtaaattg
		245I	cyacacagat	aagatactac	ccttgatgtc	ttggatagac	tgattcatct

2501 tgccatcagc atctgaaaaa atgcgcatgt tgactatttt tgcacagcgt 2551 aagccaaatc ttctgcag (SEQ ID NO:19)

SK coding sequence spans nucleotides 819-2138; coding sequence of mature peptide spans nucleotides 897-2138.

Table 10: DNA sequence of MBP*

atgaaaactg aagaaggtaa actggtaatc tggattaacg gcgataaagg ctataacggt ctcgctgaag tcggtaagaa attcgagaaa gataccggaa ttaaagtcac cgttgagcat ccggataaac tggaagagaa attcccacag 10 gttgcggcaa ctggcgatgg ccctgacatt atcttctggg cacacgaccg ctttggtggc tacgctcaat ctggcctgtt ggctgaaatc accccggaca aagcgttcca ggacaagctg tatccgttta cctgggatgc cgtacgttac aacggcaage tgattgctta cccgatcgct gttgaagcgt tatcgctgat ttataacaaa gatetgetge egaaceegee aaaaacetgg gaagagatee 15 cggcgctgga taaagaactg aaagcgaaag gtaagagcgc gctgatgttc aacctgcaag aaccgtactt cacctggccg ctgattgctg ctgacggggg ttatgcgttc aagtatgaaa acggcaagta cgacattaaa gacgtgggcg tggataacgc tggcgcgaaa gcgggtctga ccttcctggt tgacctgatt aaaaacaaac acatgaatgc agacaccgat tactccatcg cagaagctgc 20 ctttaataaa ggcgaaacag cgatgaccat caacggcccg tgggcatggt ccaacatcga caccagcaaa gtgaattatg gtgtaacggt actgccgacc ttcaagggtc aaccatccaa accgttcgtt ggcgtgctga gcgcaggtat taacgccgcc agtccgaaca aagagctggc gaaagagttc ctcgaaaact atctgctgac tgatgaaggt ctggaagcgg ttaataaaga caaaccgctg 25 ggtgccgtag cgctgaagtc ttacgaggaa gagttggcga aagatccacg tattgccgcc accatggaaa acgcccagaa aggtgaaatc atgccgaaca tecegeagat gtecgettte tggtatgeg tgegtactge ggtgateaac gccgccagcg gtcgtcagac tgtcgatgaa gccctgaaag acgcgcagac taattcgage teggtacccg geeggggate categagggt agg 30 (SEQ ID NO:20)

^{*} sequence represents cDNA sequence of MBP up to the restriction site in the polylinker where cDNA encoding SK was inserted.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: President and Fellows of Harvard College
- (ii) TITLE OF INVENTION: PLASMIN-RESISTANT STREPTOKINASE
- (iii) NUMBER OF SEQUENCES: 20
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Fish & Richardson P.C.
 - (B) STREET: 225 Franklin Street
 - (C) CITY: Boston
 - (D) STATE: MA

 - (E) COUNTRY: USA (F) ZIP: 02110-2804
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk

 - (B) COMPUTER: IBM PC compatible (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: PCT/US96/----
 - (B) FILING DATE: 07-JUN-1996
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/488,940
 - (B) FILING DATE: 09-JUN-1995
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Fraser, Janis K.
 - (B) REGISTRATION NUMBER: 34,819
 - (C) REFERENCE/DOCKET NUMBER: 05433/009W01
 - (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 617/542-5070
 - (B) TELEFAX: 617/542-8906
 - (C) TELEX: 200154
- (2) INFORMATION FOR SEQ ID NO:1:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1194 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: not relevant
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

WO 96/41883 PCT/US96/09640

- 25 -

Met Lys Thr Glu Glu Gly Lys Leu Val Ile Trp Ile Asn Gly Asp Lys Gly Tyr Asn Gly Leu Ala Glu Val Gly Lys Lys Phe Glu Lys Asp Thr Gly Ile Lys Val Thr Val Glu His Pro Asp Lys Leu Glu Glu Lys Phe Pro Gln Val Ala Ala Thr Gly Asp Gly Pro Asp Ile Ile Phe Trp Ala His Asp Arg Phe Gly Gly Tyr Ala Gln Ser Gly Leu Leu Ala Glu Ile Thr Pro Asp Lys Ala Phe Gln Asp Lys Leu Tyr Pro Phe Thr Trp Asp Ala Val Arg Tyr Asn Gly Lys Leu Ile Ala Tyr Pro Ile Ala Val Glu Ala Leu Ser Leu Ile Tyr Asn Lys Asp Leu Leu Pro Asn Pro Pro Lys Thr Trp Glu Glu Ile Pro Ala Leu Asp Lys Glu Leu Lys Ala Lys Gly Lys Ser Ala Leu Met Phe Asn Leu Gln Glu Pro Tyr Phe Thr Trp Pro Leu Ile Ala Ala Asp Gly Gly Tyr Ala Phe Lys Tyr Glu Asn Gly Lys Tyr Asp Ile Lys Asp Val Gly Val Asp Asn Ala Gly Ala Lys Ala Gly Leu Thr Phe Leu Val Asp Leu Ile Lys Asn Lys His Met Asn Ala Asp Thr Asp Tyr Ser Ile Ala Glu Ala Ala Phe Asn Lys Gly Glu Thr Ala Met Thr Ile Asn Gly Pro Trp Ala Trp Ser Asn Ile Asp Thr Ser Lys 235 Val Asn Tyr Gly Val Thr Val Leu Pro Thr Phe Lys Gly Gln Pro Ser Lys Pro Phe Val Gly Val Leu Ser Ala Gly Ile Asn Ala Ala Ser Pro Asn Lys Glu Leu Ala Lys Glu Phe Leu Glu Asn Tyr Leu Leu Thr Asp Glu Gly Leu Glu Ala Val Asn Lys Asp Lys Pro Leu Gly Ala Val Ala Leu Lys Ser Tyr Glu Glu Glu Leu Ala Lys Asp Pro Arg Ile Ala Ala 315 Thr Met Glu Asn Ala Gln Lys Gly Glu Ile Met Pro Asn Ile Pro Gln

Met Ser Ala Phe Trp Tyr Ala Val Arg Thr Ala Val Ile Asn Ala Ala Ser Gly Arg Gln Thr Val Asp Glu Ala Leu Lys Asp Ala Gln Thr Asn Ser Ser Ser Val Pro Gly Arg Gly Ser Ile Glu Gly Arg Ile Ala Gly Pro Glu Trp Leu Leu Asp Arg Pro Ser Val Asn Asn Ser Gln Leu Val Val Ser Val Ala Gly Thr Val Glu Gly Thr Asn Gln Asp Ile Ser Leu 405 410 415 Lys Phe Phe Glu Ile Asp Leu Thr Ser Arg Pro Ala His Gly Gly Lys Thr Glu Gln Gly Leu Ser Pro Lys Ser Lys Pro Phe Ala Thr Asp Ser Gly Ala Met Ser His Lys Leu Glu Lys Ala Asp Leu Leu Lys Ala Ile Gln Glu Gln Leu Ile Ala Asn Val His Ser Asn Asp Asp Tyr Phe Glu Val Ile Asp Phe Ala Ser Asp Ala Thr Ile Thr Asp Arg Asn Gly Lys Val Tyr Phe Ala Asp Lys Asp Gly Ser Val Thr Leu Pro Thr Gln Pro Val Gln Glu Phe Leu Leu Ser Gly His Val Arg Val Arg Tyr Lys Glu 520 Lys Pro Ile Gln Asn Gln Ala Lys Ser Val Asp Val Glu Tyr Thr Val Gln Phe Thr Pro Leu Asn Pro Asp Asp Asp Phe Arg Pro Gly Leu Lys Asp Thr Lys Leu Leu Lys Thr Leu Ala Ile Gly Asp Thr Ile Thr Ser Gln Glu Leu Leu Ala Gln Ala Gln Ser Ile Leu Asn Lys Asn His Pro Gly Tyr Thr Ile Tyr Glu Arg Asp Ser Ser Ile Val Thr His Asp Asn Asp Ile Phe Arg Thr Ile Leu Pro Met Asp Gln Glu Phe Thr Tyr Arg Val Lys Asn Arg Glu Gln Ala Tyr Arg Ile Asn Lys Lys Ser Gly Leu 630 Asn Glu Glu Ile Asn Asn Thr Asp Leu Ile Ser Glu Lys Tyr Tyr Val Leu Lys Lys Gly Glu Lys Pro Tyr Asp Pro Phe Asp Arg Ser His Leu

Lys Leu Phe Thr Ile Lys Tyr Val Asp Val Asp Thr Asn Glu Leu Leu Lys Ser Glu Gln Leu Leu Thr Ala Ser Glu Arg Asn Leu Asp Phe Arg Asp Leu Tyr Asp Pro Arg Asp Lys Ala Lys Leu Leu Tyr Asn Asn Leu Asp Ala Phe Gly Ile Met Asp Tyr Thr Leu Thr Gly Lys Val Glu Asp Asn His Asp Asp Thr Asn Arg Ile Ile Thr Val Tyr Met Gly Lys Arg 740 745 750 Pro Glu Gly Glu Asn Ala Ser Tyr His Leu Ala Tyr Asp Lys Asp Arg Tyr Thr Glu Glu Glu Arg Glu Val Tyr Ser Tyr Leu Arg Tyr Thr Gly Thr Pro Ile Pro Asp Asn Pro Asn Asp Lys Asn Asn Ser Gln Leu Val Val Ser Val Ala Gly Thr Val Glu Gly Thr Asn Gln Asp Ile Ser Leu Lys Phe Phe Glu Ile Asp Leu Thr Ser Arg Pro Ala His Gly Gly Lys Thr Glu Gln Gly Leu Ser Pro Lys Ser Lys Pro Phe Ala Thr Asp Ser Gly Ala Met Ser His Lys Leu Glu Lys Ala Asp Leu Leu Lys Ala Ile Gln Glu Gln Leu Ile Ala Asn Val His Ser Asn Asp Asp Tyr Phe Glu Val Ile Asp Phe Ala Ser Asp Ala Thr Ile Thr Asp Arg Asn Gly Lys 890 Val Tyr Phe Ala Asp Lys Asp Gly Ser Val Thr Leu Pro Thr Gln Pro Val Gln Glu Phe Leu Leu Ser Gly His Val Arg Val Arg Tyr Lys Glu Lys Pro Ile Gln Asn Gln Ala Lys Ser Val Asp Val Glu Tyr Thr Val 935 Gln Phe Thr Pro Leu Asn Pro Asp Asp Asp Phe Arg Pro Gly Leu Lys Asp Thr Lys Leu Leu Lys Thr Leu Ala Ile Gly Asp Thr Ile Thr Ser 965 Gln Glu Leu Leu Ala Gln Ala Gln Ser Ile Leu Asn Lys Asn His Pro Gly Tyr Thr Ile Tyr Glu Arg Asp Ser Ser Ile Val Thr His Asp Asn

WO 96/41883 PCT/US96/09640

- 28 -

Asp Ile Phe Arg Thr Ile Leu Pro Met Asp Gln Glu Phe Thr Tyr Arg 1015

Val Lys Asn Arg Glu Gln Ala Tyr Arg Ile Asn Lys Lys Ser Gly Leu 1030 1035 1040

Asn Glu Glu Ile Asn Asn Thr Asp Leu Ile Ser Glu Lys Tyr Tyr Val 1045 1050

Leu Lys Lys Gly Glu Lys Pro Tyr Asp Pro Phe Asp Arg Ser His Leu 1065

Lys Leu Phe Thr Ile Lys Tyr Val Asp Val Asp Thr Asn Glu Leu Leu 1080 1085

Lys Ser Glu Gln Leu Leu Thr Ala Ser Glu Arg Asn Leu Asp Phe Arg 1090 1095

Asp Leu Tyr Asp Pro Arg Asp Lys Ala Lys Leu Leu Tyr Asn Asn Leu 1110

Asp Ala Phe Gly Ile Met Asp Tyr Thr Leu Thr Gly Lys Val Glu Asp 1125 1130

Asn His Asp Asp Thr Asn Arg Ile Ile Thr Val Tyr Met Gly Lys Arg 1145 1140

Pro Glu Gly Glu Asn Ala Ser Tyr His Leu Ala Tyr Asp Lys Asp Arg 1160

Tyr Thr Glu Glu Glu Arg Glu Val Tyr Ser Tyr Leu Arg Tyr Thr Gly 1175 1180

Thr Pro Ile Pro Asp Asn Pro Asn Asp Lys 1185 1190

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1181 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: not relevant
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Lys Thr Glu Glu Gly Lys Leu Val Ile Trp Ile Asn Gly Asp Lys

Gly Tyr Asn Gly Leu Ala Glu Val Gly Lys Lys Phe Glu Lys Asp Thr

Gly Ile Lys Val Thr Val Glu His Pro Asp Lys Leu Glu Glu Lys Phe

Pro Gln Val Ala Ala Thr Gly Asp Gly Pro Asp Ile Ile Phe Trp Ala 55

His Asp Arg Phe Gly Gly Tyr Ala Gln Ser Gly Leu Leu Ala Glu Ile 65 70 75 80 Thr Pro Asp Lys Ala Phe Gln Asp Lys Leu Tyr Pro Phe Thr Trp Asp 85 90 95 Ala Val Arg Tyr Asn Gly Lys Leu Ile Ala Tyr Pro Ile Ala Val Glu Ala Leu Ser Leu Ile Tyr Asn Lys Asp Leu Leu Pro Asn Pro Pro Lys Thr Trp Glu Glu Ile Pro Ala Leu Asp Lys Glu Leu Lys Ala Lys Gly Lys Ser Ala Leu Met Phe Asn Leu Gln Glu Pro Tyr Phe Thr Trp Pro Leu Ile Ala Ala Asp Gly Gly Tyr Ala Phe Lys Tyr Glu Asn Gly Lys Tyr Asp Ile Lys Asp Val Gly Val Asp Asn Ala Gly Ala Lys Ala Gly Leu Thr Phe Leu Val Asp Leu Ile Lys Asn Lys His Met Asn Ala Asp 200 Thr Asp Tyr Ser Ile Ala Glu Ala Ala Phe Asn Lys Gly Glu Thr Ala Met Thr Ile Asn Gly Pro Trp Ala Trp Ser Asn Ile Asp Thr Ser Lys Val Asn Tyr Gly Val Thr Val Leu Pro Thr Phe Lys Gly Gln Pro Ser Lys Pro Phe Val Gly Val Leu Ser Ala Gly Ile Asn Ala Ala Ser Pro Asn Lys Glu Leu Ala Lys Glu Phe Leu Glu Asn Tyr Leu Leu Thr Asp Glu Gly Leu Glu Ala Val Asn Lys Asp Lys Pro Leu Gly Ala Val Ala 295 Leu Lys Ser Tyr Glu Glu Glu Leu Ala Lys Asp Pro Arg Ile Ala Ala Thr Met Glu Asn Ala Gln Lys Gly Glu Ile Met Pro Asn Ile Pro Gln Met Ser Ala Phe Trp Tyr Ala Val Arg Thr Ala Val Ile Asn Ala Ala Ser Gly Arg Gln Thr Val Asp Glu Ala Leu Lys Asp Ala Gln Thr Asn Ser Ser Val Pro Gly Arg Gly Ser Ile Glu Gly Arg Asn Asn Ser Gin Leu Val Val Ser Val Ala Gly Thr Val Glu Gly Thr Asn Gln Asp 390

Ile Ser Leu Lys Phe Phe Glu Ile Asp Leu Thr Ser Arg Pro Ala His Gly Gly Lys Thr Glu Gln Gly Leu Ser Pro Lys Ser Lys Pro Phe Ala 425 Thr Asp Ser Gly Ala Met Ser His Lys Leu Glu Lys Ala Asp Leu Leu 440 Lys Ala Ile Gln Glu Gln Leu Ile Ala Asn Val His Ser Asn Asp Asp Tyr Phe Glu Val Ile Asp Phe Ala Ser Asp Ala Thr Ile Thr Asp Arg Asn Gly Lys Val Tyr Phe Ala Asp Lys Asp Gly Ser Val Thr Leu Pro 485 Thr Gln Pro Val Gln Glu Phe Leu Leu Ser Gly His Val Arg Val Arg 505 Tyr Lys Glu Lys Pro Ile Gln Asn Gln Ala Lys Ser Val Asp Val Glu Tyr Thr Val Gln Phe Thr Pro Leu Asn Pro Asp Asp Phe Arg Pro 535 Gly Leu Lys Asp Thr Lys Leu Leu Lys Thr Leu Ala Ile Gly Asp Thr Ile Thr Ser Gln Glu Leu Leu Ala Gln Ala Gln Ser Ile Leu Asn Lys Asn His Pro Gly Tyr Thr Ile Tyr Glu Arg Asp Ser Ser Ile Val Thr 585 His Asp Asn Asp Ile Phe Arg Thr Ile Leu Pro Met Asp Gln Glu Phe 600 Thr Tyr Arg Val Lys Asn Arg Glu Gln Ala Tyr Arg Ile Asn Lys Lys Ser Gly Leu Asn Glu Glu Ile Asn Asn Thr Asp Leu Ile Ser Glu Lys Tyr Tyr Val Leu Lys Lys Gly Glu Lys Pro Tyr Asp Pro Phe Asp Arg Ser His Leu Lys Leu Phe Thr Ile Lys Tyr Val Asp Val Asp Thr Asn Glu Leu Leu Lys Ser Glu Gln Leu Leu Thr Ala Ser Glu Arg Asn Leu 680 Asp Phe Arg Asp Leu Tyr Asp Pro Arg Asp Lys Ala Lys Leu Leu Tyr 695 Asn Asn Leu Asp Ala Phe Gly Ile Met Asp Tyr Thr Leu Thr Gly Lys Val Glu Asp Asn His Asp Asp Thr Asn Arg Ile Ile Thr Val Tyr Met 730

Gly Lys Arg Pro Glu Gly Glu Asn Ala Ser Tyr His Leu Ala Tyr Asp Lys Asp Arg Tyr Thr Glu Glu Glu Arg Glu Val Tyr Ser Tyr Leu Arg Tyr Thr Gly Thr Pro Ile Pro Asp Asn Pro Asn Asp Lys Asn Asn Ser Gln Leu Val Val Ser Val Ala Gly Thr Val Glu Gly Thr Asn Gln Asp Ile Ser Leu Lys Phe Phe Glu Ile Asp Leu Thr Ser Arg Pro Ala His Gly Gly Lys Thr Glu Gln Gly Leu Ser Pro Lys Ser Lys Pro Phe Ala 825 Thr Asp Ser Gly Ala Met Ser His Lys Leu Glu Lys Ala Asp Leu Leu Lys Ala Ile Gln Glu Gln Leu Ile Ala Asn Val His Ser Asn Asp Asp 855 Tyr Phe Glu Val Ile Asp Phe Ala Ser Asp Ala Thr Ile Thr Asp Arg 870 Asn Gly Lys Val Tyr Phe Ala Asp Lys Asp Gly Ser Val Thr Leu Pro Thr Gln Pro Val Gln Glu Phe Leu Leu Ser Gly His Val Arg Val Arg Tyr Lys Glu Lys Pro Ile Gln Asn Gln Ala Lys Ser Val Asp Val Glu 920 Tyr Thr Val Gln Phe Thr Pro Leu Asn Pro Asp Asp Asp Phe Arg Pro Gly Leu Lys Asp Thr Lys Leu Leu Lys Thr Leu Ala Ile Gly Asp Thr Ile Thr Ser Gln Glu Leu Leu Ala Gln Ala Gln Ser Ile Leu Asn Lys 970 Asn His Pro Gly Tyr Thr Ile Tyr Glu Arg Asp Ser Ser Ile Val Thr His Asp Asn Asp Ile Phe Arg Thr Ile Leu Pro Met Asp Gln Glu Phe 1000 Thr Tyr Arg Val Lys Asn Arg Glu Gln Ala Tyr Arg Ile Asn Lys Lys 1015 1020 Ser Gly Leu Asn Glu Glu Ile Asn Asn Thr Asp Leu Ile Ser Glu Lys 1030 1035 Tyr Tyr Val Leu Lys Lys Gly Glu Lys Pro Tyr Asp Pro Phe Asp Arg 1045 1050 Ser His Leu Lys Leu Phe Thr Ile Lys Tyr Val Asp Val Asp Thr Asn 1060 1065

Glu Leu Lys Ser Glu Gln Leu Leu Thr Ala Ser Glu Arg Asn Leu 1075 1080 1085

Asp Phe Arg Asp Leu Tyr Asp Pro Arg Asp Lys Ala Lys Leu Leu Tyr 1090 1095 1100

Asn Asn Leu Asp Ala Phe Gly Ile Met Asp Tyr Thr Leu Thr Gly Lys 1105 1110 1115 1120

Val Glu Asp Asn His Asp Asp Thr Asn Arg Ile Ile Thr Val Tyr Met 1125 1130 1135

Gly Lys Arg Pro Glu Gly Glu Asn Ala Ser Tyr His Leu Ala Tyr Asp 1140 1150

Lys Asp Arg Tyr Thr Glu Glu Glu Arg Glu Val Tyr Ser Tyr Leu Arg 1155 1160 1165

Tyr Thr Gly Thr Pro Ile Pro Asp Asn Pro Asn Asp Lys 1170 1175 1180

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 813 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: not relevant
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Ile Ala Gly Pro Glu Trp Leu Leu Asp Arg Pro Ser Val Asn Asn Ser 1 5 10 15

Gln Leu Val Val Ser Val Ala Gly Thr Val Glu Gly Thr Asn Gln Asp 20 25 30

Ile Ser Leu Lys Phe Phe Glu Ile Asp Leu Thr Ser Arg Pro Ala His 35 40 45

Gly Gly Lys Thr Glu Gln Gly Leu Ser Pro Lys Ser Lys Pro Phe Ala 50 55 60

Thr Asp Ser Gly Ala Met Ser His Lys Leu Glu Lys Ala Asp Leu Leu 65 70 75 80

Lys Ala Ile Gln Glu Gln Leu Ile Ala Asn Val His Ser Asn Asp Asp 85 90 95

Tyr Phe Glu Val Ile Asp Phe Ala Ser Asp Ala Thr Ile Thr Asp Arg 100 105 110

Asn Gly Lys Val Tyr Phe Ala Asp Lys Asp Gly Ser Val Thr Leu Pro 115 120 125

Thr Gln Pro Val Gln Glu Phe Leu Leu Ser Gly His Val Arg Val Arg 130 135 140

Tyr Lys Glu Lys Pro Ile Gln Asn Gln Ala Lys Ser Val Asp Val Glu Tyr Thr Val Gln Phe Thr Pro Leu Asn Pro Asp Asp Phe Arg Pro 170 Gly Leu Lys Asp Thr Lys Leu Leu Lys Thr Leu Ala Ile Gly Asp Thr 180 185 Ile Thr Ser Gln Glu Leu Leu Ala Gln Ala Gln Ser Ile Leu Asn Lys Asn His Pro Gly Tyr Thr Ile Tyr Glu Arg Asp Ser Ser Ile Val Thr His Asp Asn Asp Ile Phe Arg Thr Ile Leu Pro Met Asp Gln Glu Phe 230 Thr Tyr Arg Val Lys Asn Arg Glu Gln Ala Tyr Arg Ile Asn Lys Lys Ser Gly Leu Asn Glu Glu Ile Asn Asn Thr Asp Leu Ile Ser Glu Lys 265 Tyr Tyr Val Leu Lys Lys Gly Glu Lys Pro Tyr Asp Pro Phe Asp Arg 280 Ser His Leu Lys Leu Phe Thr Ile Lys Tyr Val Asp Val Asp Thr Asn Glu Leu Leu Lys Ser Glu Gln Leu Leu Thr Ala Ser Glu Arg Asn Leu Asp Phe Arg Asp Leu Tyr Asp Pro Arg Asp Lys Ala Lys Leu Leu Tyr 325 Asn Asn Leu Asp Ala Phe Gly Ile Met Asp Tyr Thr Leu Thr Gly Lys Val Glu Asp Asn His Asp Asp Thr Asn Arg Ile Ile Thr Val Tyr Met Gly Lys Arg Pro Glu Gly Glu Asn Ala Ser Tyr His Leu Ala Tyr Asp 375 Lys Asp Arg Tyr Thr Glu Glu Glu Arg Glu Val Tyr Ser Tyr Leu Arg 395 Tyr Thr Gly Thr Pro Ile Pro Asp Asn Pro Asn Asp Lys Asn Asn Ser Gln Leu Val Val Ser Val Ala Gly Thr Val Glu Gly Thr Asn Gln Asp 425 Ile Ser Leu Lys Phe Phe Glu Ile Asp Leu Thr Ser Arg Pro Ala His Gly Gly Lys Thr Glu Gln Gly Leu Ser Pro Lys Ser Lys Pro Phe Ala Thr Asp Ser Gly Ala Met Ser His Lys Leu Glu Lys Ala Asp Leu Leu 475

Lys Ala Ile Gln Glu Gln Leu Ile Ala Asn Val His Ser Asn Asp Asp Tyr Phe Glu Val Ile Asp Phe Ala Ser Asp Ala Thr Ile Thr Asp Arg Asn Gly Lys Val Tyr Phe Ala Asp Lys Asp Gly Ser Val Thr Leu Pro 520 Thr Gln Pro Val Gln Glu Phe Leu Leu Ser Gly His Val Arg Val Arg Tyr Lys Glu Lys Pro Ile Gln Asn Gln Ala Lys Ser Val Asp Val Glu Tyr Thr Val Gln Phe Thr Pro Leu Asn Pro Asp Asp Phe Arg Pro 565 Gly Leu Lys Asp Thr Lys Leu Leu Lys Thr Leu Ala Ile Gly Asp Thr 585 Ile Thr Ser Gln Glu Leu Leu Ala Gln Ala Gln Ser Ile Leu Asn Lys Asn His Pro Gly Tyr Thr Ile Tyr Glu Arg Asp Ser Ser Ile Val Thr 615 His Asp Asn Asp Ile Phe Arg Thr Ile Leu Pro Met Asp Gln Glu Phe Thr Tyr Arg Val Lys Asn Arg Glu Gln Ala Tyr Arg Ile Asn Lys Lys 650 Ser Gly Leu Asn Glu Glu Ile Asn Asn Thr Asp Leu Ile Ser Glu Lys 665 Tyr Tyr Val Leu Lys Lys Gly Glu Lys Pro Tyr Asp Pro Phe Asp Arg Ser His Leu Lys Leu Phe Thr Ile Lys Tyr Val Asp Val Asp Thr Asn Glu Leu Leu Lys Ser Glu Gln Leu Leu Thr Ala Ser Glu Arg Asn Leu 715 Asp Phe Arg Asp Leu Tyr Asp Pro Arg Asp Lys Ala Lys Leu Leu Tyr Asn Asn Leu Asp Ala Phe Gly Ile Met Asp Tyr Thr Leu Thr Gly Lys Val Glu Asp Asn His Asp Asp Thr Asn Arg Ile Ile Thr Val Tyr Met 760 Gly Lys Arg Pro Glu Gly Glu Asn Ala Ser Tyr His Leu Ala Tyr Asp Lys Asp Arg Tyr Thr Glu Glu Glu Arg Glu Val Tyr Ser Tyr Leu Arg Tyr Thr Gly Thr Pro Ile Pro Asp Asn Pro Asn Asp Lys 805

- (2) INFORMATION FOR SEQ ID NO:4:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 800 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: not relevant
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:
 - Asn Asn Ser Gln Leu Val Val Ser Val Ala Gly Thr Val Glu Gly Thr
 1 5 10 15
 - Asn Gln Asp Ile Ser Leu Lys Phe Phe Glu Ile Asp Leu Thr Ser Arg
 20 25 30
 - Pro Ala His Gly Gly Lys Thr Glu Gln Gly Leu Ser Pro Lys Ser Lys 35 40 45
 - Pro Phe Ala Thr Asp Ser Gly Ala Met Ser His Lys Leu Glu Lys Ala 50 55 60
 - Asp Leu Leu Lys Ala Ile Gln Glu Gln Leu Ile Ala Asn Val His Ser 65 70 75 80
 - Asn Asp Asp Tyr Phe Glu Val Ile Asp Phe Ala Ser Asp Ala Thr Ile 85 90 95
 - Thr Asp Arg Asn Gly Lys Val Tyr Phe Ala Asp Lys Asp Gly Ser Val 100 105 110
 - Thr Leu Pro Thr Gln Pro Val Gln Glu Phe Leu Leu Ser Gly His Val
 - Arg Val Arg Tyr Lys Glu Lys Pro Ile Gln Asn Gln Ala Lys Ser Val
 - Asp Val Glu Tyr Thr Val Gln Phe Thr Pro Leu Asn Pro Asp Asp Asp 145 150 155 160
 - Phe Arg Pro Gly Leu Lys Asp Thr Lys Leu Leu Lys Thr Leu Ala Ile 165 170 175
 - Gly Asp Thr Ile Thr Ser Gln Glu Leu Leu Ala Gln Ala Gln Ser Ile 180 185 190
 - Leu Asn Lys Asn His Pro Gly Tyr Thr Ile Tyr Glu Arg Asp Ser Ser 195 200 205
 - Ile Val Thr His Asp Asn Asp Ile Phe Arg Thr Ile Leu Pro Met Asp 210 215 220
 - Gln Glu Phe Thr Tyr Arg Val Lys Asn Arg Glu Gln Ala Tyr Arg Ile 225 230 235 240
 - Asn Lys Lys Ser Gly Leu Asn Glu Glu Ile Asn Asn Thr Asp Leu Ile 245 250 255

Ser Glu Lys Tyr Tyr Val Leu Lys Lys Gly Glu Lys Pro Tyr Asp Pro Phe Asp Arg Ser His Leu Lys Leu Phe Thr Ile Lys Tyr Val Asp Val 280 Asp Thr Asn Glu Leu Leu Lys Ser Glu Gln Leu Leu Thr Ala Ser Glu 295 Arg Asn Leu Asp Phe Arg Asp Leu Tyr Asp Pro Arg Asp Lys Ala Lys Leu Leu Tyr Asn Asn Leu Asp Ala Phe Gly Ile Met Asp Tyr Thr Leu 325 330 335 Thr Gly Lys Val Glu Asp Asn His Asp Asp Thr Asn Arg Ile Ile Thr Val Tyr Met Gly Lys Arg Pro Glu Gly Glu Asn Ala Ser Tyr His Leu Ala Tyr Asp Lys Asp Arg Tyr Thr Glu Glu Glu Arg Glu Val Tyr Ser Tyr Leu Arg Tyr Thr Gly Thr Pro Ile Pro Asp Asn Pro Asn Asp Lys 385 390 395 400 Asn Asn Ser Gln Leu Val Val Ser Val Ala Gly Thr Val Glu Gly Thr Asn Gln Asp Ile Ser Leu Lys Phe Phe Glu Ile Asp Leu Thr Ser Arg 425 Pro Ala His Gly Gly Lys Thr Glu Gln Gly Leu Ser Pro Lys Ser Lys Pro Phe Ala Thr Asp Ser Gly Ala Met Ser His Lys Leu Glu Lys Ala Asp Leu Leu Lys Ala Ile Gln Glu Gln Leu Ile Ala Asn Val His Ser Asn Asp Asp Tyr Phe Glu Val Ile Asp Phe Ala Ser Asp Ala Thr Ile Thr Asp Arg Asn Gly Lys Val Tyr Phe Ala Asp Lys Asp Gly Ser Val 505 Thr Leu Pro Thr Gln Pro Val Gln Glu Phe Leu Leu Ser Gly His Val Arg Val Arg Tyr Lys Glu Lys Pro Ile Gln Asn Gln Ala Lys Ser Val 535 Asp Val Glu Tyr Thr Val Gln Phe Thr Pro Leu Asn Pro Asp Asp Asp 555 Phe Arg Pro Gly Leu Lys Asp Thr Lys Leu Leu Lys Thr Leu Ala Ile Gly Asp Thr Ile Thr Ser Gln Glu Leu Leu Ala Gln Ala Gln Ser Ile 585

25

Leu Asn Lys Asn His Pro Gly Tyr Thr Ile Tyr Glu Arg Asp Ser Ser 595 Ile Val Thr His Asp Asn Asp Ile Phe Arg Thr Ile Leu Pro Met Asp Gln Glu Phe Thr Tyr Arg Val Lys Asn Arg Glu Gln Ala Tyr Arg Ile Asn Lys Lys Ser Gly Leu Asn Glu Glu Ile Asn Asn Thr Asp Leu Ile Ser Glu Lys Tyr Tyr Val Leu Lys Lys Gly Glu Lys Pro Tyr Asp Pro Phe Asp Arg Ser His Leu Lys Leu Phe Thr Ile Lys Tyr Val Asp Val Asp Thr Asn Glu Leu Leu Lys Ser Glu Gln Leu Leu Thr Ala Ser Glu Arg Asn Leu Asp Phe Arg Asp Leu Tyr Asp Pro Arg Asp Lys Ala Lys Leu Leu Tyr Asn Asn Leu Asp Ala Phe Gly Ile Met Asp Tyr Thr Leu Thr Gly Lys Val Glu Asp Asn His Asp Asp Thr Asn Arg Ile Ile Thr Val Tyr Met Gly Lys Arg Pro Glu Gly Glu Asn Ala Ser Tyr His Leu Ala Tyr Asp Lys Asp Arg Tyr Thr Glu Glu Glu Arg Glu Val Tyr Ser 770 780 Tyr Leu Arg Tyr Thr Gly Thr Pro Ile Pro Asp Asn Pro Asn Asp Lys

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 25 base pairs

 - (B) TYPE: nucleic acid (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GCTGCTAGAC GCGCCATCTG TCAAC

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 23 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single

(b) ToPoLogi: Timear	
(ii) MOLECULE TYPE: cDNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:	
TGGCGCGTCT AGCAGCCACT CAG	23
(2) INFORMATION FOR SEQ ID NO:7:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 31 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: cDNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:	
CAAGACATTA GTCTGGCCTT TTTTGAAATC G	31
(2) INFORMATION FOR SEQ ID NO:8:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 23 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(ii) MOLECULE TYPE: cDNA	
(with Grovensen and control of the c	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:	
GGCCAGACTA ATGTCTTGAT TCG	23
(2) INFORMATION FOR SEQ ID NO:9:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 27 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: cDNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:	
CGATCTAACA TCGGCGCCTG CTCATGG	27

(2) INFORMATION FOR SEQ ID NO:10:

wo	96/4	1883

CGA	TCTAACA TCGGCGCCTG CTCATGG	27
(2)	INFORMATION FOR SEQ ID NO:10:	
	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 21 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:	
CGC	CGATGTT AGATCGATTT C	21
(2)	INFORMATION FOR SEQ ID NO:11:	
	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:	
GCT	CATGGAG GCGCCACAGA GGGC	24
(2)	INFORMATION FOR SEQ ID NO:12:	
	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 20 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:	
GCC	SCCTCCA TGAGCAGGTC	20
(2)	INFORMATION FOR SEQ ID NO:13:	
	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 28 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	/::\ MOT POUT P MYDEDNA	

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:	
GCTTAAGTCC GGCCTCAAAA CCATTTGC	28
(2) INFORMATION FOR SEQ ID NO:14:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(ii) MOLECULE TYPE: cDNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:	
TGAGGCCGGA CTTAAGCCTT GCTC	24
(2) INFORMATION FOR SEQ ID NO:15:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 28 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(ii) MOLECULE TYPE: CDNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:	
GCCGATCGAT ATACCGAAGA AGAACGAG	28
(2) INFORMATION FOR SEQ ID NO:16:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 31 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(ii) MOLECULE TYPE: cDNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:	
PATCGATCGG CATCATAGGC TAAATGATAG C	31
2) INFORMATION FOR SEQ ID NO:17:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1194 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: not relevant 	

- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Met Lys Thr Glu Glu Gly Lys Leu Val Ile Trp Ile Asn Gly Asp Lys

1 10 15

Gly Tyr Asn Gly Leu Ala Glu Val Gly Lys Lys Phe Glu Lys Asp Thr 20 25 30

Gly Ile Lys Val Thr Val Glu His Pro Asp Lys Leu Glu Glu Lys Phe 35 40 45

Pro Gln Val Ala Ala Thr Gly Asp Gly Pro Asp Ile Ile Phe Trp Ala 50 55 60

His Asp Arg Phe Gly Gly Tyr Ala Gln Ser Gly Leu Leu Ala Glu Ile 65 70 75 80

Thr Pro Asp Lys Ala Phe Gln Asp Lys Leu Tyr Pro Phe Thr Trp Asp 85 90 95

Ala Val Arg Tyr Asn Gly Lys Leu Ile Ala Tyr Pro Ile Ala Val Glu 100 105 110

Ala Leu Ser Leu Ile Tyr Asn Lys Asp Leu Leu Pro Asn Pro Pro Lys 115 120 125

Thr Trp Glu Glu Ile Pro Ala Leu Asp Lys Glu Leu Lys Ala Lys Gly 130 135 140

Lys Ser Ala Leu Met Phe Asn Leu Gln Glu Pro Tyr Phe Thr Trp Pro 145 150 155 160

Leu Ile Ala Ala Asp Gly Gly Tyr Ala Phe Lys Tyr Glu Asn Gly Lys 165 170 175

Tyr Asp Ile Lys Asp Val Gly Val Asp Asn Ala Gly Ala Lys Ala Gly 180 185 190

Leu Thr Phe Leu Val Asp Leu Ile Lys Asn Lys His Met Asn Ala Asp 195 200 205

Thr Asp Tyr Ser Ile Ala Glu Ala Ala Phe Asn Lys Gly Glu Thr Ala 210 215 220

Met Thr Ile Asn Gly Pro Trp Ala Trp Ser Asn Ile Asp Thr Ser Lys 235 230 235

Val Asn Tyr Gly Val Thr Val Leu Pro Thr Phe Lys Gly Gln Pro Ser 245 250 255

Lys Pro Phe Val Gly Val Leu Ser Ala Gly Ile Asn Ala Ala Ser Pro 260 265 270

Asn Lys Glu Leu Ala Lys Glu Phe Leu Glu Asn Tyr Leu Leu Thr Asp 275 280 285

Glu Gly Leu Glu Ala Val Asn Lys Asp Lys Pro Leu Gly Ala Val Ala 295 Leu Lys Ser Tyr Glu Glu Glu Leu Ala Lys Asp Pro Arg Ile Ala Ala 310 Thr Met Glu Asn Ala Gln Lys Gly Glu Ile Met Pro Asn Ile Pro Gln 330 Met Ser Ala Phe Trp Tyr Ala Val Arg Thr Ala Val Ile Asn Ala Ala Ser Gly Arg Gln Thr Val Asp Glu Ala Leu Lys Asp Ala Gln Thr Asn Ser Ser Ser Val Pro Gly Arg Gly Ser Ile Glu Gly Arg Ile Ala Gly 375 Pro Glu Trp Leu Leu Asp Ala Pro Ser Val Asn Asn Ser Gln Leu Val 395 Val Ser Val Ala Gly Thr Val Glu Gly Thr Asn Gln Asp Ile Ser Leu 410 Ala Phe Phe Glu Ile Asp Leu Thr Ser Ala Pro Ala His Gly Gly Ala 420 425 Thr Glu Gln Gly Leu Ser Pro Ala Ser Lys Pro Phe Ala Thr Asp Ser Gly Ala Met Ser His Lys Leu Glu Lys Ala Asp Leu Leu Lys Ala Ile Gln Glu Gln Leu Ile Ala Asn Val His Ser Asn Asp Asp Tyr Phe Glu Val Ile Asp Phe Ala Ser Asp Ala Thr Ile Thr Asp Arg Asn Gly Lys Val Tyr Phe Ala Asp Lys Asp Gly Ser Val Thr Leu Pro Thr Gln Pro Val Gln Glu Phe Leu Leu Ser Gly His Val Arg Val Arg Tyr Lys Glu 520 Lys Pro Ile Gln Asn Gln Ala Lys Ser Val Asp Val Glu Tyr Thr Val Gln Phe Thr Pro Leu Asn Pro Asp Asp Asp Phe Arg Pro Gly Leu Lys Asp Thr Lys Leu Leu Lys Thr Leu Ala Ile Gly Asp Thr Ile Thr Ser 565 Gln Glu Leu Leu Ala Gln Ala Gln Ser Ile Leu Asn Lys Asn His Pro 585 Gly Tyr Thr Ile Tyr Glu Arg Asp Ser Ser Ile Val Thr His Asp Asn Asp Ile Phe Arg Thr Ile Leu Pro Met Asp Gln Glu Phe Thr Tyr Arg 615 620

Val Lys Asn Arg Glu Gln Ala Tyr Arg Ile Asn Lys Lys Ser Gly Leu Asn Glu Glu Ile Asn Asn Thr Asp Leu Ile Ser Glu Lys Tyr Tyr Val 650 Leu Lys Lys Gly Glu Lys Pro Tyr Asp Pro Phe Asp Arg Ser His Leu 665 Lys Leu Phe Thr Ile Lys Tyr Val Asp Val Asp Thr Asn Glu Leu Leu Lys Ser Glu Gln Leu Leu Thr Ala Ser Glu Arg Asn Leu Asp Phe Arg Asp Leu Tyr Asp Pro Arg Asp Lys Ala Lys Leu Leu Tyr Asn Asn Leu 710 Asp Ala Phe Gly Ile Met Asp Tyr Thr Leu Thr Gly Lys Val Glu Asp Asn His Asp Asp Thr Asn Arg Ile Ile Thr Val Tyr Met Gly Lys Arg 745 Pro Glu Gly Glu Asn Ala Ser Tyr His Leu Ala Tyr Asp Lys Asp Arg Tyr Thr Glu Glu Glu Arg Glu Val Tyr Ser Tyr Leu Arg Tyr Thr Gly Thr Pro Ile Pro Asp Asn Pro Asn Asp Lys Asn Asn Ser Gln Leu Val 795 Val Ser Val Ala Gly Thr Val Glu Gly Thr Asn Gln Asp Ile Ser Leu 805 Lys Phe Phe Glu Ile Asp Leu Thr Ser Arg Pro Ala His Gly Gly Lys 825 Thr Glu Gln Gly Leu Ser Pro Lys Ser Lys Pro Phe Ala Thr Asp Ser Gly Ala Met Ser His Lys Leu Glu Lys Ala Asp Leu Leu Lys Ala Ile 855 Gln Glu Gln Leu Ile Ala Asn Val His Ser Asn Asp Asp Tyr Phe Glu 875 Val Ile Asp Phe Ala Ser Asp Ala Thr Ile Thr Asp Arg Asn Gly Lys 890 Val Tyr Phe Ala Asp Lys Asp Gly Ser Val Thr Leu Pro Thr Gln Pro 905 Val Gln Glu Phe Leu Leu Ser Gly His Val Arg Val Arg Tyr Lys Glu Lys Pro Ile Gln Asn Gln Ala Lys Ser Val Asp Val Glu Tyr Thr Val Gln Phe Thr Pro Leu Asn Pro Asp Asp Phe Arg Pro Gly Leu Lys 955

- Asp Thr Lys Leu Leu Lys Thr Leu Ala Ile Gly Asp Thr Ile Thr Ser 965 970 975
- Gln Glu Leu Leu Ala Gln Ala Gln Ser Ile Leu Asn Lys Asn His Pro 980 985 990
- Gly Tyr Thr Ile Tyr Glu Arg Asp Ser Ser Ile Val Thr His Asp Asn 995 1000 1005
- Asp Ile Phe Arg Thr Ile Leu Pro Met Asp Glu Phe Thr Tyr Arg 1010 1015 1020
- Val Lys Asn Arg Glu Gln Ala Tyr Arg Ile Asn Lys Lys Ser Gly Leu 1025 1030 1035 1040
- Asn Glu Glu Ile Asn Asn Thr Asp Leu Ile Ser Glu Lys Tyr Tyr Val 1045 1050 1055
- Leu Lys Lys Gly Glu Lys Pro Tyr Asp Pro Phe Asp Arg Ser His Leu 1060 1065 1070
- Lys Leu Phe Thr Ile Lys Tyr Val Asp Val Asp Thr Asn Glu Leu Leu 1075 1080 1085
- Lys Ser Glu Gln Leu Leu Thr Ala Ser Glu Arg Asn Leu Asp Phe Arg 1090 1095 1100
- Asp Leu Tyr Asp Pro Arg Asp Lys Ala Lys Leu Leu Tyr Asn Asn Leu 1105 1110 1115 1120
- Asp Ala Phe Gly Ile Met Asp Tyr Thr Leu Thr Gly Lys Val Glu Asp 1125 1130 1135
- Asn His Asp Asp Thr Asn Arg Ile Ile Thr Val Tyr Met Gly Lys Arg 1140 1145 1150
- Pro Glu Gly Glu Asn Ala Ser Tyr His Leu Ala Tyr Asp Lys Asp Arg 1155 1160 1165
- Tyr Thr Glu Glu Glu Arg Glu Val Tyr Ser Tyr Leu Arg Tyr Thr Gly 1170 1175 1180
- Thr Pro Ile Pro Asp Asn Pro Asn Asp Lys 1185
- (2) INFORMATION FOR SEQ ID NO:18:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1194 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: not relevant
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:
 - Met Lys Thr Glu Glu Gly Lys Leu Val Ile Trp Ile Asn Gly Asp Lys
 1 5 10 15

Gly Tyr Asn Gly Leu Ala Glu Val Gly Lys Lys Phe Glu Lys Asp Thr Gly Ile Lys Val Thr Val Glu His Pro Asp Lys Leu Glu Glu Lys Phe Pro Gln Val Ala Ala Thr Gly Asp Gly Pro Asp Ile Ile Phe Trp Ala His Asp Arg Phe Gly Gly Tyr Ala Gln Ser Gly Leu Leu Ala Glu Ile 65 70 75 80 Thr Pro Asp Lys Ala Phe Gln Asp Lys Leu Tyr Pro Phe Thr Trp Asp 85 90 95 Ala Val Arg Tyr Asn Gly Lys Leu Ile Ala Tyr Pro Ile Ala Val Glu 105 Ala Leu Ser Leu Ile Tyr Asn Lys Asp Leu Leu Pro Asn Pro Pro Lys Thr Trp Glu Glu Ile Pro Ala Leu Asp Lys Glu Leu Lys Ala Lys Gly Lys Ser Ala Leu Met Phe Asn Leu Gln Glu Pro Tyr Phe Thr Trp Pro 150 Leu Ile Ala Ala Asp Gly Gly Tyr Ala Phe Lys Tyr Glu Asn Gly Lys Tyr Asp Ile Lys Asp Val Gly Val Asp Asn Ala Gly Ala Lys Ala Gly Leu Thr Phe Leu Val Asp Leu Ile Lys Asn Lys His Met Asn Ala Asp 200 Thr Asp Tyr Ser Ile Ala Glu Ala Ala Phe Asn Lys Gly Glu Thr Ala Met Thr Ile Asn Gly Pro Trp Ala Trp Ser Asn Ile Asp Thr Ser Lys Val Asn Tyr Gly Val Thr Val Leu Pro Thr Phe Lys Gly Gln Pro Ser 250 Lys Pro Phe Val Gly Val Leu Ser Ala Gly Ile Asn Ala Ser Pro Asn Lys Glu Leu Ala Lys Glu Phe Leu Glu Asn Tyr Leu Leu Thr Asp 280 Glu Gly Leu Glu Ala Val Asn Lys Asp Lys Pro Leu Gly Ala Val Ala Leu Lys Ser Tyr Glu Glu Glu Leu Ala Lys Asp Pro Arg Ile Ala Ala Thr Met Glu Asn Ala Gln Lys Gly Glu Ile Met Pro Asn Ile Pro Gln Met Ser Ala Phe Trp Tyr Ala Val Arg Thr Ala Val Ile Asn Ala Ala

Ser Gly Arg Gln Thr Val Asp Glu Ala Leu Lys Asp Ala Gln Thr Asn Ser Ser Ser Val Pro Gly Arg Gly Ser Ile Glu Gly Arg Ile Ala Gly Pro Glu Trp Leu Leu Asp Ala Pro Ser Val Asn Asn Ser Gln Leu Val 395 Val Ser Val Ala Gly Thr Val Glu Gly Thr Asn Gln Asp Ile Ser Leu Ala Phe Phe Glu Ile Asp Leu Thr Ser Ala Pro Ala His Gly Gly Ala 425 Thr Glu Gln Gly Leu Ser Pro Ala Ser Lys Pro Phe Ala Thr Asp Ser 440 Gly Ala Met Ser His Lys Leu Glu Lys Ala Asp Leu Leu Lys Ala Ile Gln Glu Gln Leu Ile Ala Asn Val His Ser Asn Asp Asp Tyr Phe Glu 475 Val Ile Asp Phe Ala Ser Asp Ala Thr Ile Thr Asp Arg Asn Gly Lys 485 490 Val Tyr Phe Ala Asp Lys Asp Gly Ser Val Thr Leu Pro Thr Gln Pro 505 Val Gln Glu Phe Leu Leu Ser Gly His Val Arg Val Arg Tyr Lys Glu Lys Pro Ile Gln Asn Gln Ala Lys Ser Val Asp Val Glu Tyr Thr Val 535 Gln Phe Thr Pro Leu Asn Pro Asp Asp Phe Arg Pro Gly Leu Lys 555 Asp Thr Lys Leu Leu Lys Thr Leu Ala Ile Gly Asp Thr Ile Thr Ser Gln Glu Leu Leu Ala Gln Ala Gln Ser Ile Leu Asn Lys Asn His Pro 585 Gly Tyr Thr Ile Tyr Glu Arg Asp Ser Ser Ile Val Thr His Asp Asn 600 Asp Ile Phe Arg Thr Ile Leu Pro Met Asp Gln Glu Phe Thr Tyr Arg Val Lys Asn Arg Glu Gln Ala Tyr Arg Ile Asn Lys Lys Ser Gly Leu 630 Asn Glu Glu Ile Asn Asn Thr Asp Leu Ile Ser Glu Lys Tyr Tyr Val Leu Lys Lys Gly Glu Lys Pro Tyr Asp Pro Phe Asp Arg Ser His Leu Lys Leu Phe Thr Ile Lys Tyr Val Asp Val Asp Thr Asn Glu Leu Leu 680

Lys Ser Glu Gln Leu Leu Thr Ala Ser Glu Arg Asn Leu Asp Phe Arg Asp Leu Tyr Asp Pro Arg Asp Lys Ala Lys Leu Leu Tyr Asn Asn Leu Asp Ala Phe Gly Ile Met Asp Tyr Thr Leu Thr Gly Lys Val Glu Asp Asn His Asp Asp Thr Asn Arg Ile Ile Thr Val Tyr Met Gly Lys Arg Pro Glu Gly Glu Asn Ala Ser Tyr His Leu Ala Tyr Asp Ala Asp Arg Tyr Thr Glu Glu Glu Arg Glu Val Tyr Ser Tyr Leu Arg Tyr Thr Gly Thr Pro Ile Pro Asp Asn Pro Asn Asp Lys Asn Asn Ser Gln Leu Val 795 Val Ser Val Ala Gly Thr Val Glu Gly Thr Asn Gln Asp Ile Ser Leu 810 Lys Phe Phe Glu Ile Asp Leu Thr Ser Arg Pro Ala His Gly Gly Lys Thr Glu Gln Gly Leu Ser Pro Lys Ser Lys Pro Phe Ala Thr Asp Ser Gly Ala Met Ser His Lys Leu Glu Lys Ala Asp Leu Leu Lys Ala Ile Gln Glu Gln Leu Ile Ala Asn Val His Ser Asn Asp Asp Tyr Phe Glu 870 Val Ile Asp Phe Ala Ser Asp Ala Thr Ile Thr Asp Arg Asn Gly Lys 890 Val Tyr Phe Ala Asp Lys Asp Gly Ser Val Thr Leu Pro Thr Gln Pro Val Gln Glu Phe Leu Leu Ser Gly His Val Arg Val Arg Tyr Lys Glu 920 Lys Pro Ile Gln Asn Gln Ala Lys Ser Val Asp Val Glu Tyr Thr Val Gln Phe Thr Pro Leu Asn Pro Asp Asp Asp Phe Arg Pro Gly Leu Lys Asp Thr Lys Leu Leu Lys Thr Leu Ala Ile Gly Asp Thr Ile Thr Ser 970 Gln Glu Leu Leu Ala Gln Ala Gln Ser Ile Leu Asn Lys Asn His Pro Gly Tyr Thr Ile Tyr Glu Arg Asp Ser Ser Ile Val Thr His Asp Asn Asp Ile Phe Arg Thr Ile Leu Pro Met Asp Gln Glu Phe Thr Tyr Arg 1020

- Val Lys Asn Arg Glu Gln Ala Tyr Arg Ile Asn Lys Lys Ser Gly Leu 1025
- Asn Glu Glu Ile Asn Asn Thr Asp Leu Ile Ser Glu Lys Tyr Tyr Val 1045 1050
- Leu Lys Lys Gly Glu Lys Pro Tyr Asp Pro Phe Asp Arg Ser His Leu 1060 1065
- Lys Leu Phe Thr Ile Lys Tyr Val Asp Val Asp Thr Asn Glu Leu Leu
- Lys Ser Glu Gln Leu Leu Thr Ala Ser Glu Arg Asn Leu Asp Phe Arg 1100
- Asp Leu Tyr Asp Pro Arg Asp Lys Ala Lys Leu Leu Tyr Asn Asn Leu 1105 1110
- Asp Ala Phe Gly Ile Met Asp Tyr Thr Leu Thr Gly Lys Val Glu Asp 1130
- Asn His Asp Asp Thr Asn Arg Ile Ile Thr Val Tyr Met Gly Lys Arg 1145
- Pro Glu Gly Glu Asn Ala Ser Tyr His Leu Ala Tyr Asp Lys Asp Arg 1155 1160
- Tyr Thr Glu Glu Glu Arg Glu Val Tyr Ser Tyr Leu Arg Tyr Thr Gly 1175 1180

Thr Pro Ile Pro Asp Asn Pro Asn Asp Lys 1190

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2566 base pairs (B) TYPE: nucleic acid

 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CTGCAGCTAC CTGATACCAG GCATTTCCAA CAAACATGGT TAAGGCCAAA CCAAAATCAC TTTCTAGCGT TGGCAAGAGA CCTTCAAGCG AGCGCAAGAC CTTTATTGAA GTTGCTTGTC 120 GACATAAAAA TGCTGTTTGG GTTGTGCTGA TAGGCAAAAT GACCTCAAGC CCTGCAATCA 180 TCTGCTGGAG CAACTCAACT AAGTCAGCTG GTAAAACCTG CTGATGATTG AGGTAAATAA 240 ACTGAGAAGT CTCAAACAGC TGAGGGGGAT TGCCCTGATG ATCAAGCAAA TACCGCTGCC 300 AAGGTGACCC TAGCGGCTGC AAGACCTCAT ATTGACCCAA CCCCACCTCA AGTAATAAGC 360 GCTCTTTTTC GGATAAACAT GATTTGGGAA AATGCACATA TTGGTCCCCT TCTTTGACAC 420 TCACCCACTC TTTATCTCCT AACGGATGAG GGCCTACTTG CATCTCTGGA AAATAGTCTT 480

TTAGCTCCAT	AGCCATTCCT	TTCATGACGO	TCTTTAAAC	C ATTATAACAC	ATGACTCTTT	540
ATCACACAGI	TCAGTTTGTT	GTCAGCACGA	A TTTTGTATT	TCTGCCTTT1	TAATCATTAA	600
AACTAAATAA	GGGTTATTCA	TTTTTAGCA	GAACATTCA	TTAAATAGC1	ATTTATCGGA	660
ATATTAATTI	ATGTTTATGO	TAAAAAAGG	ATTATTTAC	TTTTTTCATT	GTCATTAAAA	720
TATCATTTTA	AAAAAATCAA	TAGGTTTTT	TTTGTGTCTT	TAAAACCATI	ATGTTATTCT	780
AATAATGGGG	ATTGAAACTT	AACTTTTAGG	AGGTTTCTAT	GAAAAATTAC	TTATCTTTTG	840
GGATGTTTGC	ACTGCTGTTT	GCACTAACAT	TTGGAACAGI	CAATTCTGTC	CAAGCTATTG	900
CTGGACCTGA	GTGGCTGCTA	GACCGTCCAT	CTGTCAACAA	CAGCCAATTA	GTTGTTAGCG	960
TTGCTGGTAC	TGTTGAGGGG	ACGAATCAAG	ACATTAGTCT	TAAATTTTTT	GAAATCGATC	1020
TAACATCACG	ACCTGCTCAT	AGGAAAGACA	GAGCAAGGCT	TAAGTCCAAA	ATCAAAACCA	1080
TTTGCTACTG	ATAGTGGCGC	GATGTCACAT	AAACTTGAGA	AAGCTGACTT	ACTAAAGGCT	1140
ATTCAAGAAC	AATTGATCGC	TAACGTCCAC	AGTAACGACG	ACTACTTTGA	GGTCATTGAT	1200
TTTGCAAGCG	ATGCAACCAT	TACTGATCGA	AACGGCAAGG	TCTACTTTGC	TGACAAAGAT	1260
GGTTCGGTAA	CCTTGCCGAC	CCAACCTGTC	CAAGAATTTT	TGCTAAGCGG	ACATGTGCGC	1320
GTTAGACCAT	ATAAAGAAAA	ACCAATACAA	AACCAAGCGA	AATCTGTTGA	TGTGGAATAT	1380
ACTGTACAGT	TTACTCCCTT	AAACCCTGAT	GACGATTTCA	GACCAGGTCT	CAAAGATACT	1440
AAGCTATTGA	AAACACTAGC	TATCGGTGAC	ACCATCACAT	CTCAAGAATT	ACTAGCTCAA	1500
GCACAAAGCA	TTTTAAACAA	AAACCACCCA	GGCTATACGA	TTTATGAACG	TGACTCCTCA	1560
ATCGTCACTC	ATGACAATGA	CATTTTCCGT	ACGATTTTAC	CAATGGATCA	AGAGTTTACT	1620
TACCGTGTTA	AAAATCGGGA	ACAAGCTTAT	AGGATCAATA	AAAAATCTGG	TCTGAATGAA	1680
GAAATAAACA	ACACTGACCT	GATCTCTGAG	AAATATTACG	TCCTTAAAAA	AGGGGAAAAG	1740
CCGTATGATC	CCTTTGATCG	CAGTCACTTG	AAACTGTTCA	CCATCAAATA	CGTTGATGTC	1800
GATACCAACG	AATTGCTAAA	AAGTGAGCAG	CTCTTAACAG	CTAGCGAACG	TAACTTAGAC	1860
TTCAGAGATT	TATACGATCC	TCGTGATAAG	GCTAAACTAC	TCTACAACAA	TCTCGATGCT	1920
TTTGGTATTA	TGGACTATAC	CTTAACTGGA	AAAGTAGAGG	ATAATCACGA	TGACACCAAC	1980
CGTATCATAA	CCGTTTATAT	GGGCAAGCGA	CCCGAAGGAG	AGAATGCTAG	CTATCATTTA	2040
GCCTATGATA	AAGATCGTTA	TACCGAAGAA	GAACGAGAAG	TTTACAGCTA	CCTGCGTTAT	2100
ACAGGGACAC	CTATACCTGA	TAACCCTAAC	GACAAATAAC	CACGGTCTTC	TAAAACGATG	2160
AGATTAACTG	ACAAAAAAAG	CAAGCAACAT	GCTATCAACA	GTTGCTTGCT	TTTTTCTAAC	2220
CTCTTAGTTG	TAGAGACTAG	TGACATTTCG	TGTCTAAAAT	AATCGTAACT	GGTCCATCAT	2280
TGATGAGACT	AACCTGCATA	TCTGCCCCAA	AAACGCCACG	CTCAACTGGC	ACAAAATCTG	2340
CCAATTGTTC	ATTAAAGCGA	TCATAAAACT	GGCTAGCCAT	ATCAGCTTTG	CAGCTCCTGT	2400

WO 96/41883 PCT/US96/09640

- 50 -

AAAGGCTGGG	CGATTTCCCT	TTTTGGTGTC	AGCATAAAGG	GTAAATTGCG	ACACAGATAA	2460
GATACTACCC	TTGATGTCTT	GGATAGACTG	ATTCATCTTG	CCATCAGCAT	CTGAAAAAAT	2520
GCGCATGTTG	ACTATTTTTG	CACAGCGTAA	GCCAAATCTT	CTGCAG		2566

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1143 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: double (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

ATGAAAACTG	AAGAAGGTAA	ACTGGTAATC	TGGATTAACG	GCGATAAAGG	CTATAACGGT	60
CTCGCTGAAG	TCGGTAAGAA	ATTCGAGAAA	GATACCGGAA	TTAAAGTCAC	CGTTGAGCAT	120
CCGGATAAAC	TGGAAGAGAA	ATTCCCACAG	GTTGCGGCAA	CTGGCGATGG	CCCTGACATT	180
ATCTTCTGGG	CACACGACCG	CTTTGGTGGC	TACGCTCAAT	CTGGCCTGTT	GGCTGAAATC	240
ACCCCGGACA	AAGCGTTCCA	GGACAAGCTG	TATCCGTTTA	CCTGGGATGC	CGTACGTTAC	300
AACGGCAAGC	TGATTGCTTA	CCCGATCGCT	GTTGAAGCGT	TATCGCTGAT	TTATAACAAA	360
GATCTGCTGC	CGAACCCGCC	AAAAACCTGG	GAAGAGATCC	CGGCGCTGGA	TAAAGAACTG	420
AAAGCGAAAG	GTAAGAGCGC	GCTGATGTTC	AACCTGCAAG	AACCGTACTT	CACCTGGCCG	480
CTGATTGCTG	CTGACGGGGG	TTATGCGTTC	AAGTATGAAA	ACGGCAAGTA	CGACATTAAA	540
GACGTGGGCG	TGGATAACGC	TGGCGCGAAA	GCGGGTCTGA	CCTTCCTGGT	TGACCTGATT	600
AAAAACAAAC	ACATGAATGC	AGACACCGAT	TACTCCATCG	CAGAAGCTGC	СТТТААТААА	660
GGCGAAACAG	CGATGACCAT	CAACGGCCCG	TGGGCATGGT	CCAACATCGA	CACCAGCAAA	720
GTGAATTATG	GTGTAACGGT	ACTGCCGACC	TTCAAGGGTC	AACCATCCAA	ACCGTTCGTT	780
GGCGTGCTGA	GCGCAGGTAT	TAACGCCGCC	AGTCCGAACA	AAGAGCTGGC	GAAAGAGTTC	840
CTCGAAAACT	ATCTGCTGAC	TGATGAAGGT	CTGGAAGCGG	TTAATAAAGA	CAAACCGCTG	900
GGTGCCGTAG	CGCTGAAGTC	TTACGAGGAA	GAGTTGGCGA	AAGATCCACG	TATTGCCGCC	960
ACCATGGAAA	ACGCCCAGAA	AGGTGAAATC	ATGCCGAACA	TCCCGCAGAT	GTCCGCTTTC	1020
TGGTATGCCG	TGCGTACTGC	GGTGATCAAC	GCCGCCAGCG	GTCGTCAGAC	TGTCGATGAA	1080
GCCCTGAAAG	ACGCGCAGAC	TAATTCGAGC	TCGGTACCCG	GCCGGGGATC	CATCGAGGGT	1140
AGG						1143

Other embodiments are within the following claims: What is claimed is:

- 1. A compound comprising (a) a plasminogenbinding fragment of streptokinase and (b) a blocking group at the amino-terminus of said fragment, wherein
 - (i) said compound is catalytically active;and
 - (ii) the rate of in vitro degradation of said compound in the presence of human plasminogen is at least 2 times slower compared to said rate of native streptokinase, wherein said rate is measured by the appearance of plasmin cleavage products as detected by immunoblotting using antistreptokinase antibodies.
- 2. The compound of claim 1, wherein said compound comprises the amino acid sequence of SEQ ID NO: 4.
- 3. The compound of claim 1, wherein said blocking group is a heterologous peptide.
- 4. The compound of claim 3, wherein said heterologous peptide comprises at least one heterologous amino acid.
- 5. The compound of claim 4, wherein said heterologous peptide is maltose binding protein.
- 6. A DNA comprising a coding sequence encoding the compound of claim 3.
- 7. A method of dissolving blood clots in a mammal, comprising administering to said mammal an effective amount of the compound of claim 1.

- 8. A plasminogen-binding fragment of streptokinase, wherein
 - (a) said fragment lacks between 1 and 24 amino-terminal amino acids;
 - (b) said fragment is catalytically active; and
 - (c) the rate of *in vitro* degradation of said fragment in the presence of human plasminogen is at least 2 times slower compared to said rate of native streptokinase, wherein said rate is measured by the appearance of plasmin cleavage products as detected by immunoblotting using antistreptokinase antibodies.
- 9. The fragment of claim 8, wherein said fragment comprises at least one mutation in a potential plasmin cleavage site, wherein said mutation renders said cleavage site resistant to cleavage by plasmin.
- 10. The fragment of claim 8, wherein said fragment consists of the amino acid sequence of (SEQ ID NO:4).
- 11. A DNA comprising a coding sequence encoding the fragment of claim 10.
- 12. A polypeptide comprising a plasminogenbinding fragment of streptokinase, wherein
 - (a) said fragment is catalytically active;and
 - (c) the rate of *in vitro* degradation of said polypeptide is at least two times slower compared to said rate of native streptokinase, wherein said rate is measured by the appearance of plasmin cleavage products as detected by immunoblotting using anti-streptokinase antibodies.

- 13. The polypeptide of claim 12, wherein said polypeptide comprises at least one mutation in a potential plasmin cleavage site, wherein said mutation renders said cleavage site resistant to cleavage by plasmin.
- 14. The polypeptide of claim 13, wherein said mutation is selected from the group consisting of R10A, K36A, R45A, K51A, K59A, K61A, K147A, K333, R232A, K257A, K298A, K309A, R234A, R363A, K386A, K372A, R388A, R394A, and R401A.
- 15. The polypeptide of claim 14, wherein said polypeptide comprises R10A, K36A, R45A, K51A and K59A (SEQ ID NO:17).
- 16. The polypeptide of claim 14, wherein said polypeptide comprises R10A, K36A, R45A, K51A, K59A and K386A (SEQ ID NO:18).
- 17. A DNA comprising a coding sequence encoding the polypeptide of claim 14.
- 18. A DNA comprising a coding sequence encoding the polypeptide of claim 15.
- 19. A method of dissolving blood clots in a mammal, comprising administering to said mammal an effective amount of the polypeptide of claim 15.

Purified rSK proteins

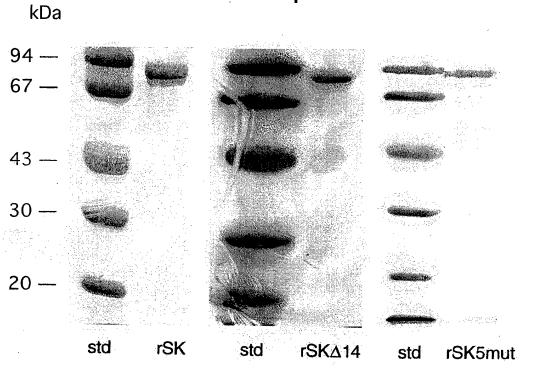


FIG. 1

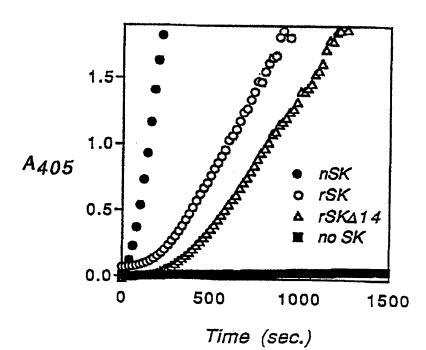
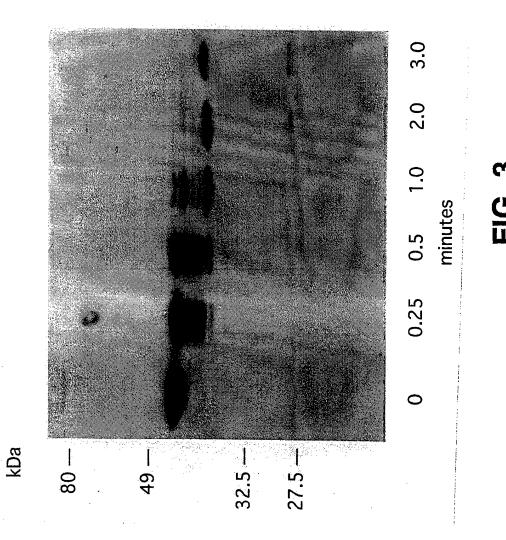


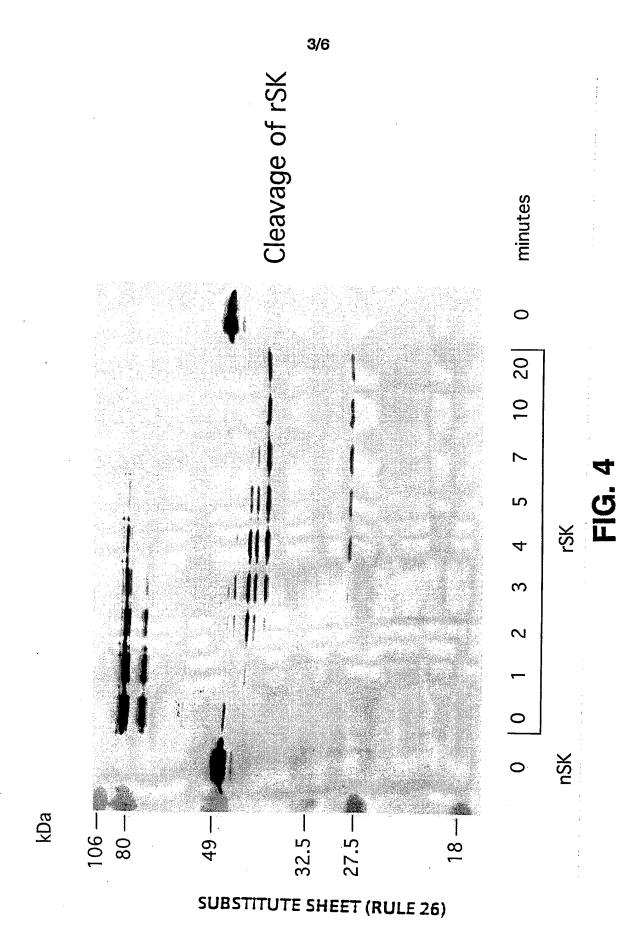
FIG. 2

SUBSTITUTE SHEET (RULE 26)

Cleavage of nSK



SUBSTITUTE SHEET (RULE 26)



4/6

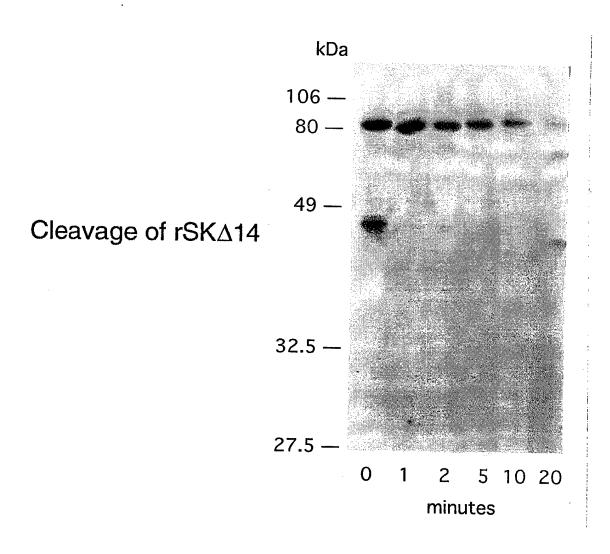
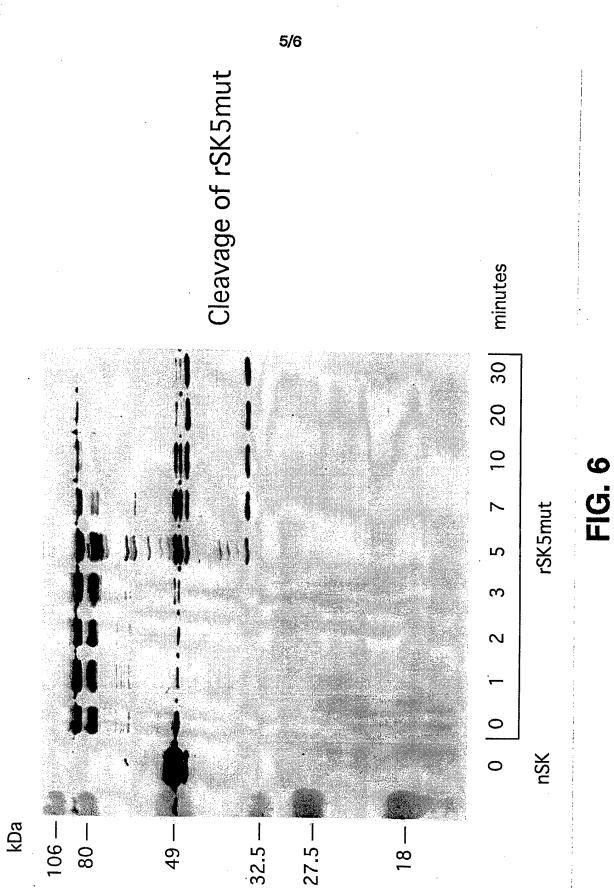
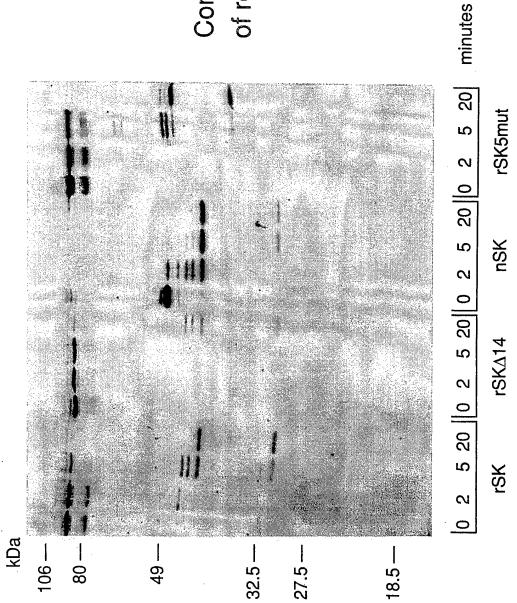


FIG. 5



SUBSTITUTE SHEET (RULE 26)





SUBSTITUTE SHEET (RULE 26)

FIG. 7

Inte onal Application No PCI/US 96/09640

A. CLASSIFICATION OF SUBJECT MATTER IPC 6 C12N15/31 C07K14/315 A61K38/16 According to International Patent Classification (IPC) or to both national classification and IPC **B. FIELDS SEARCHED** Minimum documentation searched (classification system followed by classification symbols) IPC 6 C07K Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practical, search terms used) C. DOCUMENTS CONSIDERED TO BE RELEVANT Citation of document, with indication, where appropriate, of the relevant passages Category ° Relevant to claim No. WO,A,94 07992 (GEN HOSPITAL CORP ;HARVARD COLLEGE (US)) 14 April 1994 Х 1-6,8,10-12 Y see page 3, last paragraph 9,13 see page 21, line 27 - page 24; example 2 X MOLECULAR AND GENERAL GENETICS, 1-4,6,8, vol. 212, 1988, 10-12 pages 295-300, XP002016017 C. KLESSEN ET AL: "Tripartite streptokinase gene fusion vectors for gram-positive an gram-negative procaryotes" see the whole document -/--Further documents are listed in the continuation of box C. Patent family members are listed in annex. Special categories of cited documents: "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the "A" document defining the general state of the art which is not considered to be of particular relevance invention earlier document but published on or after the international "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to filing date document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. "O" document referring to an oral disclosure, use, exhibition or document published prior to the international filing date but later than the priority date claimed in the art. "&" document member of the same patent family Date of the actual completion of the international search Date of mailing of the international search report 16 October 1996 0 4. 11. 96 Name and mailing address of the ISA Authorized officer European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Van der Schaal, C Fax: (+31-70) 340-3016

₹ 2

Inte ional Application No PC1/US 96/09640

C.(Continu	ation) DOCUMENTS CONSIDERED TO BE RELEVANT	PC170S 96/09640
Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	BIOCHEMISTRY, vol. 29, 1990, pages 3585-3590, XP002016018 D. DAVIDSON ET AL: "Plasminogen activator activities of equimolar complexes of streptokinase with variant recombinant plasminogens" see the whole document	9,13
X	US,A,5 011 686 (PANG ROY H L) 30 April 1991 see page 3; claim 11	1,3,6,7
X	WO,A,91 09125 (BRITISH BIO TECHNOLOGY) 27 June 1991 see examples 8-10	1,3,6,7
A	JOURNAL OF CLINICAL INVESTIGATION, vol. 75, no. 2, 1985, pages 413-419, XP000605367 S. RAJAGOPALAN ET AL: "A nonantigenic covalent streptokinase-polyethylene glycol complex plasminogen activator function"	
P,X	68TH SCIENTIFIC SESSION OF THE AMERICAN HEART ASSOCIATION, ANAHEIM, CALIFORNIA, USA, NOVEMBER 13-16, 1995. CIRCULATION 92 (8 SUPPL.). 1995. I623. ISSN: 0009-7322, XP002016020 LIN L-F ET AL: "Mutational studies of streptokinase identify amino acid residues critical to generation of a functional	12-14,17
Υ	SK-plasminogen activator complex." see abstract 2984	9
P,X	68TH SCIENTIFIC SESSION OF THE AMERICAN HEART ASSOCIATION, ANAHEIM, CALIFORNIA, USA, NOVEMBER 13-16, 1995. CIRCULATION 92 (8 SUPPL.). 1995. I623. ISSN: 0009-7322, XP002016019 LIU L ET AL: "Recombinant streptokinases resistant to cleavage and inactivation by	1,3,4,6,
Υ	plasmin." see abstract 2985	9,13

2

ternational application No.

PCT/US 96/09640

Box I	Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)
This Int	ernational Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:
1. X	Claims Nos.: 7,19 because they relate to subject matter not required to be searched by this Authority, namely: Remark: Although claims 7 and 19 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound.
2.	Claims Nos.: because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3.	Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).
Box II	Observations where unity of invention is lacking (Continuation of item 2 of first sheet)
This Inte	rnational Searching Authority found multiple inventions in this international application, as follows:
1.	As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2.	As all searchable claims could be searches without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3.	As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4.	No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
Remark o	The additional search fees were accompanied by the applicant's protest. No protest accompanied the payment of additional search fees.

information on patent family members

Inter ional Application No PC1/US 96/09640

	date		family ber(s)	Publication date
W0-A-9407992	14-04-94	AU-A-	5320794	26-04-94
US-A-5011686	30-04-91	NONE		
WO-A-9109125	27-06-91	US-A- AU-A- AU-B- AU-A- CA-A- CA-A- EP-A- WO-A- JP-T-	5434073 4497693 6954091 643247 6965691 2069085 2069105 0502968 0504241 9109118 5502374 5502375	18-07-95 18-11-93 18-07-91 11-11-93 18-07-91 08-06-91 08-06-91 16-09-92 23-09-92 27-06-91 28-04-93 28-04-93

Form PCT/ISA/210 (patent family annex) (July 1992)